

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:34 ; Search time 5795.58 Seconds
(without alignments)
10679.608 Million cell updates/sec

Title: US-09-930-440b-1
Perfect score: 1429
Sequence: 1 atggccttcocaaagaagaa.....ctttgaatttcattgatg 1429

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues 6940544

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:			
1:	gb_ba:		
2:	gb_btg:		
3:	gb_in:		
4:	gb_om:		
5:	gb_ov:		
6:	gb_pat:		
7:	gb_ph:		
8:	gb_pl:		
9:	gb_pr:		
10:	gb_ro:		
11:	gb_sts:		
12:	gb_sy:		
13:	gb_in:		
14:	gb_vi:		
15:	em_ba:		
16:	em_fun:		
17:	em_hum:		
18:	em_in:		
19:	em_mu:		
20:	em_om:		
21:	em_or:		
22:	em_ov:		
23:	em_ph:		
24:	em_pl:		
25:	em_pl:		
26:	em_ro:		
27:	em_sts:		
28:	em_un:		
29:	em_vi:		
30:	em_htg_hum:		
31:	em_htg_inv:		
32:	em_htg_other:		
33:	em_htg_mus:		
34:	em_htg_pln:		
35:	em_htg_rod:		
36:	em_htg_mam:		
37:	em_htg_vrt:		
38:	em_sy:		
39:	em_htgo_hum:		
40:	em_htgo_mus:		
41:	em_htgo_other:		

Pred. No. is the number of results predicted by charge to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1429	100.0	1429	6	BD270236	Human gly
2	1332.4	93.2	2536	6	AR078193	Sequence
3	1328	92.9	1552	9	AF338436	Homo sapi
4	743.8	52.1	76504	9	AL355999	Human DNA
5	658.6	46.1	960	4	SSC271330	Sus scrofa
6	606	42.4	1310	9	BC058003	Homo sapi
7	588.2	41.2	1439	10	BC022734	Mus muscu
8	433	30.3	444	6	AX119033	Sequence
9	363.6	25.4	164125	9	AC018693	Homo sapi
10	362	25.3	143506	9	AC092833	Homo sapi
11	171.2	12.0	246357	2	AC109089	Rattus no
12	170	11.9	235407	2	AC094280	Rattus no
13	157.4	11.0	198115	2	AC099712	Mus muscu
14	147.6	10.3	165444	2	AC116285	Rattus no
15	105.6	7.4	321950	1	BX294138	Pirellula
16	102	7.1	13266	1	AE013749	Yersinia
17	102	7.1	205050	1	AJ414155	Yersinia
18	90	6.3	71464	9	AL513344	Human DNA
19	90	6.3	158917	2	AC016364	Homo sapi
20	86.8	6.1	9992	1	AF004255	Vibrio ch
21	77.8	5.4	247450	1	AP005349	Vibrio vu
22	76.2	5.3	158917	2	AC016364	Homo sapi
23	74.6	5.2	302331	1	AE016810	Vibrio vu
24	70.8	5.0	198115	2	AC099712	Mus muscu
25	68.4	4.8	903	6	AR377690	Sequence
26	55.2	3.9	4987	1	AE013365	Streptoco
27	53.4	3.7	10643	1	AE006492	Streptoco
28	53.4	3.7	13338	1	AE009972	Streptoco
29	53.4	3.7	50029	1	AE014140	Streptoco
30	53.4	3.7	311600	1	AP005141	Streptoco
31	52.4	3.7	1141	6	AX083744	Sequence
32	51.4	3.6	927	6	AR375248	Sequence
33	51.4	3.6	7218	6	I66494	Sequence
34	50.2	3.5	80471	2	AC139529	Botryllus
35	49.6	3.5	110000	2	PFMAL13_3	Continuation
36	48.6	3.4	10049	1	AF291051	Carsonell
37	48.6	3.4	273275	3	AE014828	Plasmodiu
38	48.4	3.4	8056	6	AX599046	Sequence
39	47.8	3.3	161773	2	AC119335	Mus muscu
40	47.4	3.3	212664	2	AC139241	Mus muscu
41	46.8	3.3	2000	6	AX655393	Sequence
42	46.4	3.2	153751	3	AC116551	Dictyoste
43	46.2	3.2	347582	3	PFMAL4P1	Plasmodiu
44	45.6	3.2	1141	6	AX083744	Sequence
45	45.6	3.2	40068	3	CBRG44F10	Caenorhab

ALIGNMENTS

RESULT 1	BD270236	1429 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	Human glycosylation enzymes.				
DEFINITION	BD270236				
ACCESSION	BD270236.1	GI:33080004			
VERSION	JP 2002537796-A/3.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1429)				
AUTHORS	Coleman, I.A. and Betenbaugh, M.J.				
TITLE	Human glycosylation enzymes				
JOURNAL	Patent: JP 2002537796-A 3 12-NOV-2002;				

ORIGIN		/mol_type="unassigned DNA"	
Query Match		93.2%; Score 1332.4; DS 6; Length 2536;	
Best Local Similarity		94.3%; Pred. No. 0;	
Matches 1428; Conservative		0; Mismatches 1; Indels 85; Gaps 1;	
QY	1	ATGGCCCTCCCAAGAGAACTTTCAGGTCTTTGGCTGCAACCATCACGCCATGACT 60	
DB	89	ATGGCCCTCCCAAGAGAACTTTCAGGTCTTTGGCTGCAACCATCACGCCATGACT 148	
QY	61	GAGATGAGAGAACTTTCAGTATTCGTAGTACGTGGATTATCTTGTGAAGAA 120	
DB	149	GAGATGAGAGAACTTTCAGTATTCGTAGTACGTGGATTATCTTGTGAAGAA 208	
QY	121	CAGGAGTGAAGAACTTTCAGTATTCGTAGTACGTGGATTATCTTGTGAAGAA 180	
DB	209	CAGGAGTGAAGAACTTTCAGTATTCGTAGTACGTGGATTATCTTGTGAAGAA 268	
QY	181	TCAGAGCTGTCAGGTTCAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGGATCAG 240	
DB	269	TCAGAGCTGTCAGGTTCAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGGATCAG 328	
QY	241	GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTCAAGAGAACTGGCCCAATGCA 300	
DB	329	GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTCAAGAGAACTGGCCCAATGCA 368	
QY	301	GCAGAAATAGAGCTGATGGCATCGCTGATTTGCAACCGTTCTTCTCAAGCATGGACC 360	
DB	389	GCAGAAATAGAGCTGATGGCATCGCTGATTTGCAACCGTTCTTCTCAAGCATGGACC 448	
QY	361	AAGATATCTGATTAATTTCTAAGAGAGTGGCTGCGCCCTGCCCTGCCATTT 420	
DB	449	AAGATATCTGATTAATTTCTAAGAGAGTGGCTGCGCCCTGCCCTGCCATTT 508	
QY	421	TATTACATCAGATTCCTGCTTCAAGAGTGAAGATTCGCTGAGAGTGTGGAT 480	
DB	509	TATTACATCAGATTCCTGCTTCAAGAGTGAAGATTCGCTGAGAGTGTGGAT 568	
QY	481	GGGATTCGATAGATCCCAACCTTCAAGGCTGAAATTCAGTGTAGATCTCTTA 540	
DB	569	GGGATTCGATAGATCCCAACCTTCAAGGCTGAAATTCAGTGTAGATCTCTTA 628	
QY	541	GACTTCGGGCAATGTGTTGATCAGATCGCAGACAGATTTGCTTCTTTTGGGTG 600	
DB	629	GACTTCGGGCAATGTGTTGATCAGATCGCAGACAGATTTGCTTCTTTTGGGTG 688	
QY	601	GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGGAGCAGTGGG 650	
DB	689	GATGAGCAACTGTTGAGTGTCTGCTGATGGAGCAACTGGAGCAGTGGG 748	
QY	651	----- 808	
DB	749	AACTACTGGGAAAAAAGAACAAACAGATGTGGAGGCTTTTGAACAAAGACTTCTCT 808	
QY	651	----- 695	
DB	809	TTAGCCCTGAACATCAGTTTGTATCAGAGATTATCAACTTTGTCTCAAACTAGGT 868	
QY	696	TTTGAGTGTACAGACCAAGCCATCATGCTGTCTGGGATTCGAATGGGCCCA 755	
DB	869	TTTGAGTGTACAGACCAAGCCATCATGCTGTCTGGGATTCGAATGGGCCCA 928	
QY	756	CCCCGGCTTCCACTGAGAAAGCCCTCCAGGAGTGTACTGATAGTGTGAAGCTAAACTG 815	
DB	929	CCCCGGCTTCCACTGAGAAAGCCCTCCAGGAGTGTACTGATAGTGTGAAGCTAAACTG 988	
QY	816	AAGAGCTGGATTTCTTTCTTCTTCTCACTGATTTAAAGGATGGAACTTGGAGCTGTAGC 875	
DB	989	AAGAGCTGGATTTCTTTCTTCTTCTCACTGATTTAAAGGATGGAACTTGGAGCTGTAGC 1048	
QY	876	TAGTGCCTCTCTATCAAAATCAGGGTTTGCACTTGGAGACATAATCTACCTTAAATAGTGC 935	

1049	TAGTGCCTCTCTATCAAAATCAGGGTTTGCACTTGGAGACATAATCTACCTTAAATAGTGC 1108	
QY	936	ATTTTTCCTCAGGGAATTTTAGATGAATTCGAATAAACTCTCTAGCAAAATGAATCTC 995
DB	1109	ATTTTTCCTCAGGGAATTTTAGATGAATTCGAATAAACTCTCTAGCAAAATGAATCTC 1168
QY	996	ACAAATAGCAATGAGGTACCTTTTGTGAGCCCTTAAAAAGTCTTATTTTGTGAAGGGCAA 1055
DB	1169	ACAAATAGCAATGAGGTACCTTTTGTGAGCCCTTAAAAAGTCTTATTTTGTGAAGGGCAA 1228
QY	1056	AAACTCTAGGAGTCAAACTCTCAGTATTCATTTTCAAGATTTTGTGAGGAAATTT 1115
DB	1229	AAACTCTAGGAGTCAAACTCTCAGTATTCATTTTCAAGATTTTGTGAGGAAATTT 1288
QY	1116	CTGTTTATGATGATGAATGGAATCAAGAGGAAATTTGTAATTTGATTTAATTCATCTGTC 1175
DB	1289	CTGTTTATGATGATGAATGGAATCAAGAGGAAATTTGTAATTTGATTTAATTCATCTGTC 1348
QY	1176	TTTAGGAGCTCTCAATATCTCGGTCTCTGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1235
DB	1349	TTTAGGAGCTCTCAATATCTCGGTCTCTGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1408
QY	1236	TTAAACCTACTAATATATCTCTTTCATTTTAAATAATTTTCAATTTTGAATCTAGGAAACTC 1295
DB	1409	TTAAACCTACTAATATATCTCTTTCATTTTAAATAATTTTCAATTTTGAATCTAGGAAACTC 1468
QY	1296	TGAGTCTCTGATTTAGGAGGAGCTTTAATACCAAACTGTAACATGCTCAACTGTATA 1355
DB	1469	TGAGTCTCTGATTTAGGAGGAGCTTTAATACCAAACTGTAACATGCTCAACTGTATA 1528
QY	1356	CAACTCAAAATACACAGCTCATTTGGCTGCTCAGTCTAACTCTAGAAATGGATGCTTTTG 1415
DB	1529	CAACTCAAAATACACAGCTCATTTGGCTGCTCAGTCTAACTCTAGAAATGGATGCTTTTG 1588
QY	1416	AATTCATTTCTGATG 1429
DB	1589	AATTCATTTCTGATG 1602

RESULT 3	AF338436	1552 bp	mRNA	linear	PRI 08-MAY-2001
LOCUS	Homo sapiens Clorf13 mRNA, complete cds.				
DEFINITION	AF338436				
ACCESSION	AF338436.1				
VERSION	GI:13430284				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Sood, R., Bonner, T.I., Makalowska, I., Stephan, D.A., Robbins, C.M., Connors, T.D., Morgenbesser, S.D., Su, K., Faruque, M.U., Pinkett, H., Graham, C., Baxeianis, A.D., Klinger, K.W., Landes, G.M., Trent, J.M. and Carpten, J.D.				
TITLE	Cloning and characterization of 13 novel transcripts and the human RGS8 gene from the 1q25 region encompassing the hereditary prostate cancer (HPC1) locus				
JOURNAL	Genomics 73 (2), 211-222 (2001)				
MEDLINE	21218927				
PUBMED	11318611				
REFERENCE	2 (bases 1 to 1552)				
AUTHORS	Sood, R., Bonner, T.I., Makalowska, I., Stephan, D.A., Robbins, C.M., Connors, T.D., Morgenbesser, S.D., Su, K., Faruque, M.U., Pinkett, H., Graham, C., Baxeianis, A.D., Klinger, K.W., Landes, G.M., Carpten, J.D. and Trent, J.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-JAN-2001) Laboratory of Genetics, National Institute of Mental Health, Bldg 35, Room 3D06, MSC 4094, Bethesda, MD 20892, USA				
FEATURES	Location/Qualifiers				
source	1..1552				
	/organism="Homo sapiens"				

825	TTTGGAGTGTTCACAGACCAAAAGCCATCATGCTCTGGTCTCTGGGATTCATATGGGCCCA	884
756	CCCGGCTTCCACTGCAGAAAGCCTCCAGGAGCTTTACTGATAGTGTCTGAAGCTAAACTG	815
885	CCCGGCTTCCACTGCAGAAAGCCTCCAGGAGCTTTACTGATAGTGTCTGAAGCTAAACTG	944
816	AAGAGCCTGGATTTCCCTTTCTTTCTCTGATTTTAAAGATGGAACCTTTGGAAGCTGGTAGC	875
945	AAGAGCCTGGATTTCCCTTTCTTTCTCTGATTTTAAAGATGGAACCTTTGGAAGCTGGTAGC	1004
876	TAGTGCTCTCTATCAAAATCAGAGGTTTGGACCTTGGACATATCTACCTTTAAATAGTGC	935
1005	TAGTGCTCTCTATCAAAATCAGAGGTTTGGACCTTGGACATATCTACCTTTAAATAGTGC	1064
936	ATTTTTCTCAGGGAATTTTAGATGAACCTTGAATTAACCTCTCTAGCAATGAATCTCTC	995
1065	ATTTTTCTCAGGGAATTTTAGATGAACCTTGAATTAACCTCTCTAGCAATGAATCTCTC	1124
996	ACAATAAGCAATTGAGGTACCTTTTGTGAGCCTTAAAGATCTTATTTGTGAAGGGGCAA	1055
1125	ACAATAAGCAATTGAGGTACCTTTTGTGAGCCTTAAAGATCTTATTTGTGAAGGGGCAA	1184
1056	AACTCTAGGAGTCACAACTCTCAGTCACTCAATTCACAGATTTTTTTTGTGGAGAAATTT	1115
1185	AACTCTAGGAGTCACAACTCTCAGTCACTCAATTCACAGATTTTTTTTGTGGAGAAATTT	1244
1116	CTGTTTATATGGATGAATGGAATCAAGAGGAAATTTGTAATTCATTAATTCATCTCTGTC	1175
1245	CTGTTTATATGGATGAATGGAATCAAGAGGAAATTTGTAATTCATTAATTCATCTCTGTC	1304
1176	TTTAGGAGCTCTCAATATCTCGGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1235
1305	TTTAGGAGCTCTCAATATCTCGGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1364
1236	TTAAACCACTATAATATCT	1295
1365	TTAAACCACTATAATATCT	1424
1296	TGAGTACTGCAATTTAGGAGGCACTTTAATACCACTGTAACATGCTCTCACTGTAATA	1355
1425	TGAGTACTGCAATTTAGGAGGCACTTTAATACCACTGTAACATGCTCTCACTGTAATA	1484
1356	CACTCAAAATACACCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1415
1485	CACTCAAAATACACCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1544
1416	AATTCATT 1423	
1545	AATTCATT 1552	
RESULT 4		
AL355999	Human DNA sequence from clone RP11-24906 on chromosome 1, complete	PRI 23-AUG-2001
LOCUS	75504 bp DNA linear	
DEFINITION	sequence.	
ACCESSION	AL355999.9	GI:15384805
VERSION	AL355999.9	
KEYWORDS	HTG.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	1 (bases 1 to 75504)	
AUTHORS	Tracey, A.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire, UK. E-mail enquiries: humquerry@sanger.ac.uk	
COMMENT	On Aug 31, 2001 this sequence version replaced gi:13990035. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations	

Db	825	TTTGGAGTGTACAGACCAAGCAATCATGACTCTGTGTCTCTGGATTCCTCAATGGGCCA	884
QY	756	CCCGGGTTCACATCGAGAAAGCTCCAGGAGTTTACTGATAGTGTGAAGCTAAACTG	815
Db	885	CCCGGGTTCACATCGAGAAAGCTCCAGGAGTTTACTGATAGTGTGAAGCTAAACTG	944
QY	816	AAGAGCTGATTTCTTCTTCTTCACTGATTTAAAGGATGGAACCTTGGAGCTGGTAGC	875
Db	945	AAGAGCTGATTTCTTCTTCTTCACTGATTTAAAGGATGGAACCTTGGAGCTGGTAGC	1004
QY	876	TAGTCCCTCTCTATCAAAATCAGGTTTGCACCTTGAGACATAACTACCTTAAATAGTC	935
Db	1005	TAGTCCCTCTCTATCAAAATCAGGTTTGCACCTTGAGACATAACTACCTTAAATAGTSC	1064
QY	936	ATTTTCTTCTCAGGGAATTTTAGATGAATTAATAAATCTCTCTAGCAATGAATCTC	995
Db	1065	ATTTTCTTCTCAGGGAATTTTAGATGAATTAATAAATCTCTCTAGCAATGAATCTC	1124
QY	996	ACAATAAGCATTGAGGTACCTTTTGTGAGCCTTAAAGCTTATTTTGTGAAGGGCA	1055
Db	1125	ACAATAAGCATTGAGGTACCTTTTGTGAGCCTTAAAGCTTATTTTGTGAAGGGCA	1184
QY	1056	AACTCTAGGAGTCAAACTCTCAGTCTTCAATTCACAGATTTTGTGAGAGAAATTT	1115
Db	1185	AACTCTAGGAGTCAAACTCTCAGTCTTCAATTCACAGATTTTGTGAGAGAAATTT	1244
QY	1116	CTGTTTATATGATGAATGAAATGAAATCAAGAGAGAAATTTGATTTGATTAATTCATCTGTC	1175
Db	1245	CTGTTTATATGATGAATGAAATGAAATCAAGAGAGAAATTTGATTTGATTAATTCATCTGTC	1304
QY	1176	TTTAGGAGCTCTCATTTATCTCGGCTCTGCTTCTTAAATCTTAAAGTTTCTTAAT	1235
Db	1305	TTTAGGAGCTCTCATTTATCTCGGCTCTGCTTCTTAAATCTTAAAGTTTCTTAAT	1364
QY	1236	TTAAACCACTAATATATGCTCTTCAATTTAAATAATTTCAATTTGGAATCTAGGAAACTC	1295
Db	1365	TTAAACCACTAATATATGCTCTTCAATTTAAATAATTTCAATTTGGAATCTAGGAAACTC	1424
QY	1296	TCAGTCTGCTGATTTAGGAGGAGCTTTAAATACCAAACTGTAACATGCTCAACTGATA	1355
Db	1425	TCAGTCTGCTGATTTAGGAGGAGCTTTAAATACCAAACTGTAACATGCTCAACTGATA	1484
QY	1356	CAACTCAAAATACACAGCTCAATTTGGTGTCTGCTCACTCACTGAGAGGATCTTTTG	1415
Db	1485	CAACTCAAAATACACAGCTCAATTTGGTGTCTGCTCACTCACTGAGAGGATCTTTTG	1544
QY	1416	AATTCATTT 1423	
Db	1545	AATTCATTT 1552	
RESULT 4			
AL355999 76504 bp DNA linear PRI 23-AUG-2001			
LOCUS Human DNA sequence from clone RP11-24906 on chromosome 1, complete			
DEFINITION sequence.			
ACCESSION AL355999			
VERSION AL355999.9 GI:15384805			
KEYWORDS HTG.			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
REFERENCE 1 (bases 1 to 76504)			
AUTHORS Tracey, A.			
TITLE Direct Submission			
JOURNAL Submitted (23-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,			
CS10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk			
requests: clonerrequest@sanger.ac.uk			
On Aug 31, 2001 this sequence version replaced gi:13990035.			
During sequence assembly data is compared from overlapping clones.			
Where differences are found these are annotated as variations			

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrl> Rpl1-24906 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone Rpl1-24906. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone Rpl1-24906 is at 76504 in this sequence. The true left end of clone Rpl1-333G16 is at 55978 in this sequence. The true right end of clone Rpl1-71D4 is at 2000 in this sequence.

FEATURES		Location/Qualifiers	
source			
..:..76504			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/chromosome="1"			
/clone="RP11-249O6"			
/clone_lib="RPC1-11.1"			
ORIGIN			
Query Match 52.1%; Score 743.8; DB 9; Length 76504;			
Best Local Similarity 98.4%; Pred. No. 5.2e-180;			
Matches 751; Conservative 0; Mismatches 12; Indels 0; Gaps 0;			
QY	557	GATTATCAACCTTTGTTGTCAACCTAGGTTTGGAGTGTCAAGACCAAGCCATCATGA	726
Db	20322	GTTTCTCATTTGTTTTCCTCGTCTAGGTTTGGAGTGTCAAGACCAAGCCATCATGA	20381
QY	727	CTCTGGTCTCTGGATTCGAATGGGCGCCCGGCTTCCACTGCAGAAAGCCTCCAGGG	786
Db	20382	CTCTGGTCTCTGGATTCGAATGGGCGCCCGGCTTCCACTGCAGAAAGCCTCCAGGG	20441
QY	787	AGTTTACTGATGCTGAAGCTAACTGAAGAGCCTGGATTTCTTTTCTTCTACTGATT	846
Db	20442	AGTTTACTGATGCTGAAGCTAACTGAAGAGCCTGGATTTCTTTTCTTCTACTGATT	20501
QY	847	TAAAGATGGAACCTTGGAGCTGTGAGTGTAGTGTCTCTATCAATCAGGTTTGCAC	906
Db	20502	TAAAGATGGAACCTTGGAGCTGTGAGTGTAGTGTCTCTATCAATCAGGTTTGCAC	20561
QY	907	CTTGACATATCTACCTTAATGTGATTTTCTTCTCAGGAAATTTTAGATGAACTT	966
Db	20562	CTTGACATATCTACCTTAATGTGATTTTCTCAGGAAATTTTAGATGAACTT	20621
QY	967	GAATAAATCTCTCAGCAATGAATCTCACAATAAGCAATGAGTACCTTTTGTAGCC	1026
Db	20622	GAATAAATCTCTCAGCAATGAATCTCACAATAAGCAATGAGTACCTTTTGTAGCC	20681
QY	1027	TTAAAAGCTTATTTTGTGAAGGGGCAAAATCTTAGGAGTCAAACTCTCAGTCAATTC	1086
Db	20682	TTAAAAGCTTATTTTGTGAAGGGGCAAAATCTTAGGAGTCAAACTCTCAGTCAATTC	20741

QY	1087	ATTTCACAGATTTTTCGTGGAGAAATTCCTTTATATGATGAATGAATCAAGAG 1146	
DB	20742	ATTTCACAGATTTTTCGTGGAGAAATTCCTTTATATGATGAATGAATCAAGAG 20801	
QY	1147	AAAATGTAATGATTAATTCATCTCTTTAGGAGCTCTCATTTATCTCGGTCTCTGGT 1206	
DB	20802	AAAATGTAATGATTAATTCATCTCTTTAGGAGCTCTCATTTATCTCGGTCTCTGGT 20861	
QY	1207	TCCATCTCCTATTTTAAAGTTCCTAAATTTTAAACCACTATATATGCTTCAATTTAAT 1266	
DB	20852	TCCATCTCCTATTTTAAAGTTCCTAAATTTTAAACCACTATATATGCTTCAATTTAAT 20921	
QY	1267	AATATTTCAATTTGGAACTCTAGGAAACTCTGAGCTACTGAGTACTGCAITTAGCGGACATTTAAT 1326	
DB	20922	AATATTTCAATTTGGAACTCTAGGAAACTCTGAGTACTGAGTACTGCAITTAGCGGACATTTAAT 20981	
QY	1327	ACCAAACTGTAACTGTCTCAACTGTATACAACTCAAAATACACCACTCATTGCGCTGC 1386	
DB	20982	ACCAAACTGTAACTGTCTCAACTGTATACAACTCAAAATACACCACTCATTGCGCTGC 21041	
QY	1387	TCAGTCTCACTAGAACTGATGCTTTTGAATTCATTTCTGATG 1429	
DB	21042	TCAGTCTCACTAGAACTGATGCTTTTGAATTCATTTCTGATG 21084	

RESULT 5
LOCUS SSC271330 960 bp mRNA linear MAM 01-FEB-2001
DEFINITION Sus scrofa mRNA for acylneuraminase lyase (npl gene).
ACCESSION AJ271330
VERSION AJ271330.1 GI:12657443
KEYWORDS acylneuraminase lyase; npl gene.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE 1 Bruse, P., Traving, C. and Schauer, R.
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 960)
REFERENCE Bruse, P.
AUTHORS Direct Submission
JOURNAL Submitted (14-JAN-2000) Bruse P., Biochemisches Institut,
Christian-Albrechts-Universitaet, Olshausenstrasse 40, Kiel 24098,
Germany
COMMENT Alignment with the acylneuraminase lyase cDNA sequence from pig kidney (nplcna) enables the assembly of the following mouse partial ESTs: gBA1550577, gBA162738, gBA14179, gBA544828, dbJAV158604, dbJAV170531, gBA547408, and of the following human partial ESTs: gBW79930, gBR97395, gBA2226558, gBH65991, gBA1916625, gBA1521193, gBA1079177, to complete open reading frames homologous to microbial acylneuraminase lyases.

1	REFERENCE	Bruse, P., Traving, C. and Schauer, R.
2	AUTHORS	Unpublished
3	REFERENCE	2 (bases 1 to 960)
4	AUTHORS	Bruse, P.
5	TITLE	Direct Submission
6	JOURNAL	Submitted (14-JAN-2000) Bruse P., Biochemisches Institut, Christian-Albrechts-Universitaet, Olshausenstrasse 40, Kiel 24098, Germany
7	COMMENT	Alignment with the acylneuraminate lyase cDNA sequence from pig kidney
8		(nplna) enables the assembly of the following mouse partial ESTs: gbAI550577, gbAA162738, gbW14179, gbAA544828, dbJAV158804, dbJAV170531, gbAA547408, and of the following human partial ESTs: gbW79930, gbR97395, gbAA226558, gbH65991, gbAI916625, gbAI521193, gbAI079177, to complete
9	FEATURES	open reading frames homologous to microbial acylneuraminate lyases.
10	source	Location/Qualifiers
11		1..960
12		/organism="Sus scrofa"
13		/mol_type="mRNA"
14		/db_xref="taxon:9823"
15		/tissue_type="kidney"
16		1..960
17	gene	/gene="npl"
18		1..960
19	CDS	/gene="npl"
20		/codon_start=1
21		/product="acylneuraminate lyase"
22		/protein_id="CAC27797.1"
23		/db_xref="GI:12657444"
24		/db_xref="GOA:Q9BEC7"
25		/db_xref="SPTREMBL:Q9BEC7"
26		/translation="MASPKKKLQGLVAATITPMTEHGEINFSVIGQVYDLVEQGVK NIFVNGTGBGLSLISERQVABEWVTKGNKLDQIVIHVGLSLSESQELQHAAK

IGADGIAVIAPPELKPWNKDNILNLEKEVAAAAPALPFFYYHIFALTVKIVRAEELLD
GIQKPTFFGLXSEFDLDFGQVDQNDHQRFAPLFGVDEQLLSALVNGATGAVGS
TYNLRKTNQMLEAFERKDFSSALNHQFCIQRFINFWKLGFGVSTKAIMTLVSGI
PMGPPLPLQKASREFDNAKAKLSLDVLSFDLCKGNLEAC

ORIGIN		Query Match	46.1%; Score 658.6; DB 4; Length 960;
		Best Local Similarity	83.4%; Pred. No. 2.9e-158;
		Matches 796; Conservative	0; Mismatches 74; Indels 85; Gaps 1;
QY	1	ATGSCCTTCCCAAGAAACTTTCAGGCTCTGCTGGCTGCAACCATCACCCCAATGACT	60
DB	1		
QY	61	GAGAAATCGAGAAATCACTTTTCAGTAAATGGTCAGTATGGAATATCTTTGGAAGAA	120
DB	61		
QY	121	CAGGAGCTGAGAACATTTTGTGAATGCGCAACAGAGAGAGCGCTGCTCCCTGAGCGTC	180
DB	121		
QY	181	TCAGAGCGTCCCGAGTTCAGAGAGAGTGGTGCACAAAGGAAGCAACAGCTGATCAG	240
DB	181		
QY	241	GTGATAATTCAGTAGGAGCTAGCTTGAAGAGTCAAGAGACTGCGCCCAACATGCA	300
DB	241		
QY	301	GCAGAAATAGAGCGTGGTGGATCGCTCATCTGACACCGTCTTCTTCAAGCCATGGACC	360
DB	301		
QY	361	AAAGATATCTGATTAATTTCTTAAGGAAGTGGCTGTGCGCCCTGCTCCCTGCCATTT	420
DB	361		
QY	421	TATTACTATCATTTCTGCTTGAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGT	480
DB	421		
QY	481	GGATATCTGGAATAGATCCACCTTCCAAAGGGCTGAATTCAGTGATACAGATCTCTTA	540
DB	481		
QY	541	GACTTCGGGCAATGTGTGATCAGAAATCCCGAGCAACAGTTTGCCTTCCTTTGGGGTG	600
DB	541		
QY	601	GATGAGCACTGTTGAGTGTCTGTTGATGGGAGCACTGGAGCAGTGGG-----	660
DB	601		
QY	651	-----	
DB	661	AACCTACCTGGGAGAAAGCAACACAGATTTTGGAGGCTTTTGAACGAAGGACTTCTCT	720
QY	651	-----	
DB	721	TCAGCGCTGAATCATCAGTTTTCATTCAAGAGATTTATCACTTTGTGTCAAACTAGGT	780
QY	696	TTTGGAGCTGCAGACCAAGCCATCATGCTGCTCTGGGATTCATGGGCCCA	755
DB	781	TTTGGAGCTGCAGACCAAGCCATCATGCTGCTCTGGGATTCATGGGCCCA	840
QY	756	CCCCGGCTTCCACTGACAGAAAGCTCCAGGAGTTTACTGATAGTGTCTGAAGCTAAACTG	815
DB	841	CCCCGGCTTCCACTGACAGAAAGCTCCAGGAGTTTACTGATAGTGTCTGAAGCTAAACTG	900
QY	816	AAAGAGCTGGATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT	872
DB	901	AAAGAGCTGGATTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT	957

RESULT 5
BC058003
LOCUS
DEFINITION
Homo sapiens cDNA clone MGC:61869 IMAGE:4607228, complete cds.
ACCESSION
BC058003.1 GI:34783661
VERSION
MGC.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1310)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, C.M., Scheuer, C.F., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, S.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Rulyk, S.W.,
Villalón, D.K., Muzny, D.V., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettman, M., Madan, A., Madan, A., Rodrigues, S.,
Sachez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Schnerk, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1310)
Strausberg, R.
Direct Subission
Submitted (08-SEP-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 48 Row: d Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 13540532.

FEATURES
Location/Qualifiers
1..1310
/organism="Homo sapiens"
/mol_type="mRNA"

```

/db_xref="taxon:3606"
/clone="MGC:61569 IMAGE:4607228"
/tissue_type="kidney"
/clone_lib="NIH MGC_75"
/lab_host="DH10B"
/note="Vec:or: pDNR-LIB"
109..811
CDs      misc_feature
109..811
/codon_start=1
/product="Unknown (protein for MGC:61869)"
/protein_id="AAH58003.1"
/db_xref="GI:34783662"
/translating="MA3PKK1:QGLVAATIPMTENGINFSVIGQYDYLKBOGVK
NIFPNTGGELSLVSYSERQVABEWTKGDKLQVLIHVCALSLKESQELAAHAE
IGAGIATIAFFFLKPMNDLILNFKLEVAAPALPYHYHIALTKKRAEELLD
KLDKIPFQGLKSDTDLDFGQCVQDQNRQOFALFEGWDFICLQRFNFVVKLENS
KLVSKNGRTPLGTFNTPFLH"
124..717
/nc="DapA; Region: Dihydropicolinate
synthase/N-acetylneuraminate lyase [Amino acid transport
and metabolism / Cell envelope biogenesis, outer
membrane]"
/db_xref="CDD:COG0329"

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 9.9e-145; Length 1310;
Matches 606; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGGCTTCCCAAGAAGAACTTCAGGGTCCTGTGSGCTGCAACCATCATCGCAATGACT 60
Db 109 ATGGGCTTCCCAAGAAGAACTTCAGGGTCCTGTGSGCTGCAACCATCATCGCAATGACT 168
QY 61 GAGATGAGAAATCACTTTCAGTAATGTGTGATGTGTGATGTGTGATGTGTGATGTG 120
Db 169 GAGATGAGAAATCACTTTCAGTAATGTGTGATGTGTGATGTGTGATGTGTGATGTG 228
QY 121 CAGGAGTGAAGAACATTTTGTGATGTCACACAGGAGAGGCGCTGCTTGCAGCGTC 180
Db 229 CAGGAGTGAAGAACATTTTGTGATGTCACACAGGAGAGGCGCTGCTTGCAGCGTC 288
QY 181 TCAGAGCGTCGCGCAGGTTGACAGAGAGTGGGTGACAAAAGGGAAGGACAAAGCTGATCAG 240
Db 289 TCAGAGCGTCGCGCAGGTTGACAGAGAGTGGGTGACAAAAGGGAAGGACAAAGCTGATCAG 348
QY 241 GTGATAATTCAGTAGGAGCATGAGCTTGAAGAGTCACAGGAATCGGCCCAACATGCA 300
Db 349 G-GATAAATTCAGTAGGAGCATGAGCTTGAAGAGTCACAGGAATCGGCCCAACATGCA 408
QY 301 GAGAAATAGGAGCTGATCGCATCGCTGTCATTGCAACCGTCTTCTTCAAGCCATGGACC 360
Db 409 GAGAAATAGGAGCTGATCGCATCGCTGTCATTGCAACCGTCTTCTTCAAGCCATGGACC 468
QY 361 AAAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCGCTGCGCTGCGCATTT 420
Db 469 AAAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCGCCCGCTGCGCTGCGCATTT 528
QY 421 TATTACTATCATCTCTGCTTGAAGAGTGAAGATTCGCTGAGAGTGTGTTGGAT 480
Db 529 TATTACTATCATCTCTGCTTGAAGAGTGAAGATTCGCTGAGAGTGTGTTGGAT 588
QY 481 GGGATTCGATTAAGA-TCCCACTTCCCAAGGCTGAATTCAGTGATACAGATCTCTTA 540
Db 589 GGGATTCGATTAAGATCCCACTTCCCAAGGCTGAATTCAGTGATACAGATCTCTTA 648
QY 541 GACTTCGGCAATGTGTGTATCAGATTCGACCAACAGAGTGTGCTTTCCTTTGGGGTG 600
Db 649 GACTTCGGCAATGTGTGTATCAGATTCGACCAACAGAGTGTGCTTTCCTTTGGGGTG 708
QY 601 GATGAG 606
Db 709 GATGAG 714

```

```

RESULT 7
EC022734
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
FEATURES
source
gene

```

```

EC022734      1439 bp      mRNA      linear      ROD 06-OCT-2003
Mus musculus N-acetylneuraminate pyruvate lyase, mRNA (cDNA clone
MGC:31604 IMAGE:4506776), complete cds.
EC022734
MGC.
BC022734.1 GI:18490966
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1439)
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
Hepkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L.,
Schetz T.E., Brownstein M.J., Ustin T.B., Tschiyuki S.,
Carrinci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,
Abramson R.D., Mullen J., Sosa S.A., McKwan P.J.,
McKernan K.J., Valle J.A., Gunaratne P.H., Richards S.,
Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalobos D.K., Muzny D.M., Sodergren E.C., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettner M., Madan A., Rodriguez S.,
Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,
Bouffard G.G., Blakesley R.W., Tsachman J.W., Green E.D.,
Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Bickertfield Y.S., Krzywinski M.I., Skalska U., Smalinski D.E.,
Scherch A., Schein J.E., Jones S.J. and Marra M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 15999-16903 (2002)
22388257
12477932
2 (bases 1 to 1439)
Strausberg R.
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
Series: IRK Plate: 44 Row: 2 Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21311854.
Location/Qualifiers
1..1439
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:31604 IMAGE:4506776"
/tissue_type="Eye, retina, mouse strain C57BL/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..1439
/gene="Npl"

```



```
QY 1 ATGGGCTCCCAAGAGAGAACTTCAAGGCTTTGGCTGCAACCAATCAGCCATGACT 60
Db 1 ATGGGCTCCCAAGAGAGAACTTCAAGGCTTTGGCTGCAACCAATCAGCCATGACT 60
QY 61 GAGATGGAGAAATCACTTTTCAGTAATGGTCACTATGTGATTAATCTTGTGAAGAA 120
Db 61 GAGATGGAGAAATCACTTTTCAGTAATGGTCACTATGTGATTAATCTTGTGAAGAA 120
QY 121 CAGGGAGTGAAGAACATTTTGTGAATGACACACAGGAGAGGCGCTGCTCCCTGACGGTC 180
Db 121 CAGGGAGTGAAGAACATTTTGTGAATGACACACAGGAGAGGCGCTGCTCCCTGACGGTC 180
QY 181 TCAGACGCTGCCAGCTTCAGAGAGAGTGGTGCACAAAAGGAGGAGACAGCTGGATCAG 240
Db 181 TCAGACGCTGCCAGCTTCAGAGAGAGTGGTGCACAAAAGGAGGAGACAGCTGGATCAG 240
QY 241 GTGATAATTCACGTAGAGACACTGAGCTTGAAGAGTCAAGGAACTGGCCCAACATGCA 300
Db 241 GTGATAATTCACGTAGAGACACTGAGCTTGAAGAGTCAAGGAACTGGCCCAACATGCA 300
QY 301 CGAGAAATAGGAGCTGACGGCATCGCTGATTCATTCGACCGTCTTCTCTCAAGCCATGGACC 360
Db 301 CGAGAAATAGGAGCTGACGGCATCGCTGATTCATTCGACCGTCTTCTCTCAAGCCATGGACC 360
QY 361 AAGATATCCTGATTAATTTCTTAAGGAGAGTGGCTGCTCCGCGCCCTGCCCTGCCATTT 420
Db 361 AAGATATCCTGATTAATTTCTTAAGGAGAGTGGCTGCTCCGCGCCCTGCCCTGCCATTT 420
QY 421 TATTACTATCACATTCCTGCTTGA 445
Db 420 TATTACTATCACATTCCTGCTTGA 444

RESULT 9
AC018693 164125 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-555N21 from 2, complete sequence.
DEFINITION AC018693
ACCESSION AC018693.8 GI:15321565
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164125)
AUTHORS Suistat,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PubMed 9847074
REFERENCE 2 (bases 1 to 164125)
AUTHORS Belter,E., Cotton,M. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP11-555N21
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 164125)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 164125)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 164125)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 28, 2001 this sequence version replaced gi:14141805.
```

```
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0555N21
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., paired quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Fietz de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-490C19; the clone sequenced to the right is RP11-173C1, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-555N21; actual end is at base position 31977 of RP11-173C1.

Polymorphisms exist between RP11-555N21 and RP11-173C1.

FEATURES	Location/Qualifiers
source	1. 164125 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="2" /map="2"
misc_feature	/clone="RP11-555N21" 1128. 1542 /notes="similar to Homo sapiens EST AV756546 (NID:910914394)"
misc_feature	1136. 1543 /notes="similar to Homo sapiens EST A1123298 (NID:93539064); ov48C08.x1" 1137. 1548 /notes="similar to Homo sapiens EST BF573448 (NID:911647160)"
misc_feature	1138. 1482 /notes="similar to Homo sapiens EST BF695725 (NID:911981133)"
misc_feature	1175. 1542 /notes="similar to Homo sapiens EST BG942016 (NID:914341388) ax20a10.x1"
misc_feature	1180. 1479 /notes="similar to Homo sapiens EST AV763398 (NID:910921246)"
misc_feature	1182. 1543


```
/note="similar to Homo sapiens EST AW156903 (NID:g5228304)
au89d05.x1"
misc_feature
1274..1556
/note="similar to Homo sapiens EST AA483370 (NID:g2212183)
ne74c02.s"
misc_feature
1274..1553
/note="similar to Homo sapiens EST AA614576 (NID:g2466772)
np53d10.s1"
misc_feature
1274..1542
/note="similar to Homo sapiens EST AA722249 (NID:g2739956)
zh21f04.s1"
repeat_region
1533..1553
/rpt_family="AT-rich"
repeat_region
2087..2396
/rpt_family="Alu"
repeat_region
2367..2399
/rpt_family="(A)n"
misc_feature
2373..2398
/note="similar to Homo sapiens EST BG942015
(NID:g14341388) ax20a10.x1"
misc_feature
2379..2401
/note="similar to Homo sapiens EST AV756546
(NID:g10914394)"
repeat_region
2612..2792
/rpt_family="MIR"
repeat_region
2799..2837
/rpt_family="(CTA)n"
repeat_region
3039..3105
/rpt_family="L2"
repeat_region
3168..3467
/rpt_family="Alu"
repeat_region
3604..3829
/rpt_family="MaLR"
repeat_region
3990..4294
/rpt_family="Alu"
repeat_region
4313..4403
/rpt_family="ERV1"
repeat_region
4404..4706
/rpt_family="Alu"
repeat_region
4707..4769
/rpt_family="ERV1"
misc_feature
4784..5032
/note="similar to Homo sapiens EST BF876452
(NID:g12266582)"
misc_feature
4811..5032
/note="similar to Homo sapiens EST AL525867
(NID:g12789360)"
misc_feature
4926..5032
/note="similar to Bos taurus EST BF653570 (NID:gl1318702)"
repeat_region
5038..5069
/rpt_family="(T)n"
repeat_region
5040..5351
/rpt_family="Alu"
misc_feature
5058..5076
/note="similar to Homo sapiens EST AV728536
(NID:g10837957)"
misc_feature
5164..5166
/note="similar to Homo sapiens EST BF740024
(NID:g12066700)"
misc_feature
5353..5725
/note="similar to Homo sapiens EST AL554919
(NID:g12896160)"
misc_feature
5358..5902
/note="similar to Homo sapiens EST AL577521
(NID:g12940733)"
misc_feature
5362..5687
/note="similar to Homo sapiens EST ALS25867
(NID:g12789360)"
misc_feature
5362..5498
/cote="similar to Homo sapiens EST BF876452
(NID:g12266582)"
misc_feature
5364..5856
/note="similar to Homo sapiens EST BF740024
```

```
(NID:g-2066700)"
5364..5523
/note="similar to Bos taurus EST BF653570 (NID:g-1918702)"
5457..5687
/note="similar to Homo sapiens EST H65991 (NID:g1024731)
yr73b09.r1"
misc_feature
5473..5725
/note="similar to Homo sapiens EST AW192585 (NID:g6471284)
xl47a06.x1"
misc_feature
5548..5725
/note="similar to Homo sapiens EST AV728536
(NID:g10837957)"
misc_feature
5569..5725
/note="similar to Homo sapiens EST AI245446 (NID:g3840843)
gk29f09.x1"
misc_feature
5637..5725
/note="similar to Homo sapiens EST AI916625 (NID:g5656480)

Query Match 25.4%; Score 363.6; DB 9; Length 164125;
Best Local Similarity 83.0%; Pred. No. 3.3e-82;
Matches 478; Conservative 0; Mismatches 84; Indels 14; Gaps 5;

QY 487 CTGGATAGATCCACACCTTCCAGGGCTGAAATTCAGTGATACAGATCTCTAGACTTC 546
Db 5347 CCGGCCAAGAGCCCACTTCTTAAAGGGCTGAAATTCAGTGATACAGATCTCTAGACTTC 5404
QY 547 GGGCAATGTTGATCAGATCCGCCAGCAACAGTTTGCTTCTCTTTTGGGGTGGATGAG 606
Db 5405 AGGCAATGTTGATCAGATCCGCCAGCAACAGTTTGCTTCTCTTTTGGGGTGGATGAG 5464
QY 607 CAACTGTTGAGTGCTCTGTTGATGGAGCAACTGGAGCAGTGGGAGTTTGTATCAGA 666
Db 5465 CAACTGTTGAGTGCTCTGTTGATGGAGCAACTGGAGCAGTGGGAGTTTGTATCAGA 5524
QY 667 GATTATCAACTTTGTTGTCAACTAGGTTTGGAGTGTCACAGACCAAGCCATCATGA 726
Db 5525 GATTATCAACTTTTGTAGTCAAACTAGGTTTCTAGTGTACAGATCAAAAGCCATCATGA 5584
QY 727 CTCTGTCCTCTGGATTCCAACTGGGCCACCCGGCTTCCACTGCAGAAAGCCCTCCAGGG 786
Db 5585 CTCTGTCCTCTGGAGTGGAGTGGGCCCAACCCCATCTTCGGTGCAGAAAGCCCTTAGGG 5644
QY 787 AGTTTACTGATAGTGCTGAAGCTAAACTGAAGAGCCTGGAATTCCTTTCTTCTCAGTAT 846
Db 5645 AGTTTACTGATAGTGCTGAAGCCAACTAAAGAGCTTGGATTT-TTTTCTTCTCAGTAT 5703
QY 847 TAAAGATGGAACCTTGGAGCTGGTAGTGGCTCTCTATCAATCAGGGTTTGCAC 906
Db 5704 TAAAGATGGAACCTTGGAGCCTATAGCCAGTGGCTGTCTATCCCAATCAGGCTGTGTC 5763
QY 907 CTTCAGACATATCTACCTTAAATAGTGCATTTTTTTTCTCAGGGA-ATTTAGATGAAT 965
Db 5764 ATTAAGACATAGTCTGCCCTTAATAGTGCATTTTCTCAGGGAGCTTTTGAAGAAACT 5823
QY 966 TGA---TAACTCTCTAGCAAAATGAAATCTCA-----CAATAGCATTTGAGGTACC 1015
Db 5824 TGAATCTGAATCTCTCTAGCAAAATGAAATCTCAATCTCAATCAACAGATATTTAAGTACT 5883
QY 1016 TTTTGTAGCCCTTAAAGTCTTATTTCTGAGGG 1051
Db 5884 TACTATGAGCCCTTAAAGGTTTATTTTAGGCCGGG 5919

RESULT 10
AC092833
LOCUS 143506 bp DNA linear PRI 29-MAY-2002
DEFINITION Homo sapiens BAC clone RP13-490C19 from 2, complete sequence.
ACCESSION AC092833
VERSION AC092833.4 GI:20331024
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

Mammalia; Butheria; Primates; Catarrhini; Homnidae; Homo.
1. (bases 1 to 143506)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074

REFERENCE
AUTHORS
TITLE
JOURNAL

2. (bases 1 to 143506)
Grewal,N., Haglund,K. and Boyer,E.
The sequence of Homo sapiens BAC clone RP13-490C19
Unpublished (2001)
3. (bases 1 to 143506)
Waterston,R.H.
Direct Submission
Submitted (30-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

4. (bases 1 to 143506)
Waterston,R.H.
Direct Submission
Submitted (26-APR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

5. (bases 1 to 143506)
Waterston,R.
Direct Submission
Submitted (29-MAY-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 26, 2002 this sequence version replaced gi:17647021.
----- Genome Center
Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

REFERENCE
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_FHC0490C19

cells (BRL Life Technologies). The library has been arrayed into 384-well microtiter dishes and also gridded onto 22x22cm nylon high density filters for screening by probe hybridization.

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-255N19, 2000 bp overlap.
Actual start of this clone is at base position 72571 of RP11-255N19; actual end is at base position 143506 of RP13-490C19.

Polymorphisms have been identified between AC018693 and AC092833.

Data from AC018693 was used to finish this clone, AC092833.

FEATURES		Location/Qualifiers
source	1..143506	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:3606" /chromosome="2"
repeat_region	1..171	/clone="RP13-490C19" /clone_lib="RP13-13" /map="2" /rpt_family="Alu"
repeat_region	266..397	/rpt_family="ERV1"
repeat_region	401..472	/rpt_family="L1"
repeat_region	642..931	/rpt_family="Alu"
repeat_region	942..1239	/rpt_family="Alu"
repeat_region	1530..1771	/rpt_family="Alu"
repeat_region	1793..2089	/rpt_family="Alu"
repeat_region	2180..2252	/rpt_family="Alu"
repeat_region	2367..2501	/rpt_family="Alu"
repeat_region	2503..2807	/rpt_family="Alu"
repeat_region	3049..3083	/rpt_family="Alu"
repeat_region	3219..3519	/rpt_family="Alu"
repeat_region	3647..3960	/rpt_family="Alu"
repeat_region	4133..4465	/rpt_family="Alu"
repeat_region	7329..7349	/rpt_family="AT-rich"
repeat_region	7883..8185	/rpt_family="Alu"
repeat_region	8398..8578	/rpt_family="MIR"
repeat_region	8585..8623	/rpt_family="CTA)n"
repeat_region	8825..8891	/rpt_family="L2"
repeat_region	8954..9253	/rpt_family="Alu"
repeat_region	9321..9631	/rpt_family="MaLR"
repeat_region	9776..10080	/rpt_family="Alu"
repeat_region	10099..10189	/rpt_family="ERV1"
repeat_region	10190..10492	/rpt_family="Alu"
repeat_region	10493..10554	/rpt_family="ERV1"
repeat_region	10825..11137	/rpt_family="Alu"

Db	11430	AGTTTACTGATAGTCTGAAGCCAACTAAAAAGCTTGATTT-TTTTCTTCACTGGCT	11440
Qy	847	TAAAGGATGAAACTTGGAAAGCTGGTAGCTAGTGCCTCTCTATCAAAATCAGGTTTGAC	906
Db	11489	TAAAGATGGAAGCTTGGAAAGCTTATAGCCAGTCTGTCTATCCAAATCAGGCTGTGTC	11548
Qy	907	CTTGAGACATAATCTACCTTAAATAGTGCATTTTCTCAGGGA-ATTTAGATGAAT	965
Db	11549	ATTAAGACATAGTCTGCCCTAAATAGTGCATTTTCTCAGGACTTTTATAGAAAAT	11608
Qy	966	TGAA---TAAACTCTCTAGCAAAATGAATCTCA-----CAATAAGCATTGAGTACC	1015
Db	11609	TGAATGAACTCTCTCTAGCAAAATGAATCTCACTTCAATCAACAGATATTAAAT	11668
Qy	1016	TTTTGTGAGCCTTAAAAAGTCTTTATTTGTGAAGGG	1051
Db	11669	TACTATGAGCCTTAAAAAGGTTTATTTTAGCGCGG	11704
RESULT 11			
AC109089/c			
LOCUS	Rattus norvegicus clone CH230-202E3, *** SEQUENCING IN PROGRESS		
DEFINITION	***, 4 unordered pieces.		
ACCESSION	AC109089.5		
VERSION	AC109089.5 GI:25072913		
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus		
ORGANISM	Rattus norvegicus		
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	1 (bases 1 to 246357)		
AUTHORS	Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Chen,Y., Chou,J., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Diya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Louiseged,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindratre,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathewine,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Mirja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemelehm,O., Okwunu,G., Olampunsgoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shivartsbeyn,A., Sisson,I., Sittler,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,		

AC109089	246357 bp	DNA	linear	HTG 19-NOV-2002
Rattus norvegicus clone CH230-202E3	***	SEQUENCING IN PROGRESS		
AC109089	4 unordered pieces.			
AC109089.5	GI:25072913			
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.				
Rattus norvegicus (Norway rat)				
Rattus norvegicus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
1 (bases 1 to 246357)				
Muzny, D. Marie, Metzker, M. Lee, Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Bialwal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, S., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, N., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, N., Hernandez, J., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Herrandez, J., Hollins, B., Howell, S., Hui, S., Hui, J., Idlerbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshe, L., Louised, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Maup, P., Martin, K., Martin, R., Martine, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Mirza, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okwono, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quirio, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, E., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savory, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,				

Db 119552 CCTGAAACACGAGTGTGTGTCG 119531

RESULT 12
AC094280
LOCUS
DEFINITION Rattus norvegicus clone CH230-3G22, WORKING DRAFT SEQUENCE, 3
unordered pieces.
AC094280
ACCESSION AC094280.10 GI:24954121
VERSION HTG; HTGS_FULL; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 235407)
REFERENCE
AUTHORS Murry D.Marie, Metzker M.Lee, Abramson S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,
Anyalebechi V., Ayodeji A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Ceasar H., Chen A.,
Chacko J.J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle X., Cree A., D'Souza L.,
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
DeGado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escoto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Freaser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Georgopoulos E., Geer K., Gill R., Grady M., Guerra M., Guevara W.,
Guraratne P., Haaland M., Hamil C., Hamilton C., Hamilton K.,
Harvey Y., Haykal P., Hawes A., Henderson N., Hernandez J.,
Herrandez R., Hines S., Hladun S.L., Hodgson A., Hoques M.,
Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Karpathy S., Kelly S., Kelly S., Khan Z., King D., Kowar C.,
Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorenshewa L., Lulseged H., Lozado R.J., Lu X., Ma J.,
Maheshwari M., Mahindartine M., Mahmoud M., Malloy K., Mangum A.,
Margum B., Mapua P., Martin K., Martin R., Martinez E.,
Mawney S., McLeod M.P., McNeill T.Z., Meenen E.,
Milosavljevic A., Miner G., Mirza E., Montemayor J., Moore S.,
Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
Nwakoelameh O., Okwuonu G., Olarnpunsagoon A., Pal S., Parks K.,
Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,
Plopper F., Poindexter A., Popovic D., Primus E., Pu L.,
Puzo M., Quirroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
Sanders W., Savary G., Scherer S., Scott G., Shatsman S., Shen H.,
Shetty J., Shvartsbeyn A., Sisson T., Sitter C.D., Smajs D.,
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa C.,
Steinle M., Strong R., Sutton A., Svatek A., Taber F., Taylor C.,
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,
Valas R., Vera V., Villalana D., Waldron L., Walker B., Wang J.,
Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,
Williams G., Willson R., Wleczek R., Wooden H., Worley K.,
Wright B., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Zhu D., von
Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,
Weinstock G. and Gibbs R.A.
Direct Submission
Unpublished
2 (bases 1 to 235407)
REFERENCE
AUTHORS Worley K.C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 235407)
Direct Submission
Submitted (14-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 14, 2002 this sequence version replaced gi:23266009.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAZZ
Center clone name: CH230-3G22
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 226392 bases at least Q40
Consensus quality: 228011 bases at least Q30
Consensus quality: 228972 bases at least Q20
Estimated insert size: 230729; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
* 1 232242: contig of 232242 bp in length
* 232243 232342: gap of unknown length
* 232343 233631: contig of 1289 bp in length
* 233632 233731: gap of unknown length
* 233732 235407: contig of 1676 bp in length.
----- Location/Qualifiers
1. 235407
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-3G22"
1. 1252
/note="wgs_end_extension"
clone_end:Sp6"
4523..5686
/note="wgs_end_extension"
clone_end:Sp6"
5995..7857
/note="wgs_end_extension"
clone_end:Sp6"
8387..133612
/note="clone boundary"
clone_end:Sp6
site:
end_sequence:BH31006f"
complement(229906..230730)
/note="clone_boundary"

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

Baylor Plaza, Houston, TX 77030, USA
 3 (bases - to 165444)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:22856515.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GUCR
 Center clone name: CH230-362N7
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 158039 bases at least Q40
 Consensus quality: 159792 bases at least Q30
 Consensus quality: 150671 bases at least Q20
 Estimated insert size: 161405; sum-of-contigs estimation
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Gensack_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 164095: contig of 164095 bp in length
 164096 164195: gap of unknown length
 164196 165444: contig of 1249 bp in length.

Location/Qualifiers
 1..165444
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-362N7"
 1..1192
 /note="wgs_contig"

misc_feature
 ORIGIN
 Query Match 10.3%; Score 147.6; DE 2; Length 165444;
 Best Local Similarity 68.8%; Pred. No. 1.2e-26;
 Matches 253; Conservative 0; Mismatches 99; Indels 16; Gaps 3;
 QY 632 GAGCACTGGAGCTGGGAGTTTGTATCCAGAGATTTATCACTTCTGTCAACT 691
 Db 4-607 GAACAGAAGGCTCAGTCACCAAGTCTCATCTCCAGGATTTATCACTATG 41548
 QY 692 AGCTTTGGAGTGCAGACCAAGCCATCATGACTCTGCTCTCGGATTCGAATGGG 751
 Db 4-547 AGCTTTGGAGTGCAGACCAAGGATATCATGACCTGTCTCTGGATTCGAATGGG 41488
 QY 752 CCCACCCCGGCTTCCACTGCGAAGAGCTCCAGGGAGTTTACTGATAGTCTGAAGCTAA 811

Db 41487 CCCACCCCGGCTTCCACTGCGAAGATCCACCGAGCTTTCTCTGCTAATG----- 41438
 QY 812 ACTGAAGAGCTGGATTTCTTTCTTCTCACTGATTTAAAGATGGAACCTGGAGCTGG 871
 Db 41437 -TTGAAGAGCTTGAATTTCCCTTCTTCCCTAGATTAAAGATGGGACTTGGAGACTG 41379
 QY 872 TACTAGTGGCTCTCTATCAAAATCAGGTTTGACCTTTGACACATATCTACCTTAATA 931
 Db 41378 TACTGGTGC-----CTGATAGCTAGTGTGTAGCCGAGACTTTATCTATCTTCAATA 41323
 QY 932 GTCCATTTTCTCAGGGA-ATTTAGATGAACCTTGAATAAATCTCTCTAGCAATGAA 990
 Db 41322 GTCCATTTCTCAGGACTTTTAGGTGAATTTGACCAAGCTTTCTTTAGCAGTA 41263
 QY 991 ATCTCACA 998
 Db 41262 GAATCCCA 41255

RESULT 15
 EX294138/c
 LOCUS BX294138 321950 bp DNA linear BCT 11-JUL-2003
 DEFINITION Pirellula sp. strain 1 complete genome; segment 6/24.
 ACCESSION BX294138 EX119912
 VERSION BX294138.1 GI:32443466
 KEYWORDS complete genome.
 SOURCE Pirellula sp. 1
 ORGANISM Pirellula sp. 1
 Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
 Planctomycetaceae; Pirellula.
 REFERENCE 1 (Bases 1 to 321950)
 Gloeckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T.,
 Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R.,
 Schlöter, H., Amann, R. and Reinhardt, R.,
 Complete genome sequence of the marine planctomycete Pirellula sp.
 strain 1
 Proc. Natl. Acad. Sci. U.S.A. 100 (14), 8298-8303 (2003)
 MEDLINE 2735913
 PUBMED 12835416
 REFERENCE 2 (Bases 1 to 321950)
 Kube, M., Borzym, K., Heitmann, K., Klages, S., Marguardt, I.,
 Leitz, S., Beck, A., Pawlik, R., Reinhardt, R., Gloeckner, F.O.,
 Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Rabus, R.,
 Schleutner, H. and Amann, R.
 Direct Submission
 Submitted (21-JAN-2003) Max Planck Institute for Molecular
 Genetics, proscience Innestrasse 73, D-14195 Berlin, Germany Max
 Planck Institute for Marine Microbiology Celsiusstrasse 1, D-28359
 Bremen, Germany
 This project was carried out by
 *Max Planck Institute for Molecular Genetics, Berlin, Germany; *Max
 Planck Institute for Marine Microbiology, Bremen, Germany; in the
 framework of the REGX-project, http://www.regx.de -----
 Genome Center
 Center: Max Planck Institute for Molecular Genetics
 Center code: MPIMG
 ----- Summary Statistics
 Sequencing vector: pUC19; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 7142841 bases at least Q40
 Consensus quality: 7145138 bases at least Q30
 Consensus quality: 7145484 bases at least Q20
 Quality coverage: 8.03

 This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid sequence; assembly was additionally confirmed by long
 range PCR and cosmid end sequences.

See <http://www.micro-genomes.mpg.de/pirellula/> for more information including minimal tiling path from a set of 223 cosmids out of 908. See the misc feature tag below for the boundaries of the MTP cosmids. ----- Annotation

Center: Max Planck Institute for Marine Microbiology
Celsiusstrasse 1, D-28359 Bremen, Germany.

Center Code: MEIMM
Email: fcg@mpi-bremen.de
Phone: +49 (0)421 2028 938 Fax: +49 (0)421 2028 580

Three different programs (Glimmer, Critica and Orpheus) were used for ORF-prediction. A nonredundant list of ORFs was generated by suitable parsing of the results.
Automated annotation was done with the software package Pedant Pro (<http://www.biomax.de>). All ORF predictions and annotations were manually corrected by considering all results of the different tools applied. See <http://www.regx.de> for more information and access to supplementary information.

FEATURES

source

Location/Qualifiers

```

1..321950
  /organism="Pirellula sp. 1"
  /mol_type="genomic DNA"
  /strain="1"
  /db_xref="taxon:243090"
  complement(362..1051)
  /locus_tag="RB2934"
  complement(362..1051)
  /locus_tag="RB2934"
  /function="lipid, fatty-acid and isoprenoid biosynthesis; fatty acid biosynthesis; lipid, fatty-acid and isoprenoid metabolism"
  /note="PMID: 2144277 PMID: 10784042 PMID: 11481430 best DB hits: BLAST: swissprot:O67611; ACP_AQUAE ACYL CARRIER PROTEIN (ACP) ----; E=0.018 swissprot:P19372; ACP_RHIME ACYL CARRIER PROTEIN (ACP) ----; E=0.055 swissprot:O9RG22; ACP_RHILE ACYL CARRIER PROTEIN (ACP) ----; E=0.064 COG: aq_17174; COG0236 Acyl carrier protein; E=0.002 PFAM: PF00950; Phosphopantetheine attachment site; E=0.3"
  /codon_start=1
  /transl_table=11
  /product="similar to acyl carrier protein"
  /protein_id="CAD72909.1"
  /db_xref="GI:32443467"
  /translation="MGLDAVEIMVIEVDEHFGISINDETRVLTWGLVALIQSRIGA AEIATCPITLSRSLRSVRELTNDQTLRIIGTRVDMNRTQERQWQDLDLIGT AAPGLRPAILKRLAFLATTFVIAFLGSVTIDVAILPLTIALAACATLALHTIP FRSIPDAAATGCAIARRMAGISVATKQLHLRTDSEILLHELPRIVAAATLGTGSKITR TTRFEDLGMG"
  complement(1184..1408)
  /locus_tag="RB2936"
  complement(1184..1408)
  /locus_tag="RB2936"
  /function="cellular communication/signal transduction"
  /note="PMID: 8134129 best DB hits: BLAST: swissprot:PS3356; HT16 HYDAT TYROSINE-PROTEIN KINASE HTK16 ----; E=0.007 gb:AAF37305.1; AC005931_3 (AC005931) 6-Phosphofructo-2-kinase; E=0.008 gb:AAF58766.1; AE003828) CG16728 gene product [Drosophila; E=0.009 PFAM: PF00023; Ank repeat; E=2.4e-08"
  /codon_start=1
  /transl_table=11
  /product="similar to tyrosine-protein kinase HTK16"
  /protein_id="CAD72910.1"
  /db_xref="GI:32443468"
  /translation="MQAAGLHPRRTFGVSPSANMLTQLMDHHRHTPLHEACTQGTDR TVRALLEYGADRVYLDANNAQTAEASAASHN"
  complement(1416..1907)
  /locus_tag="RB2939"
  complement(1416..1907)
  /locus_tag="RB2939"
  /note="PMID: 8943436 best DB hits: BLAST: embi:CAB66271.1;

```

```

(AL136519) putative integral-membrane protein.; E=le-06
emb1:CAB72202.1; (AL138851) putative integral-membrane protein; E=5e-06"
/codon_start=1
/transl_table=11
/product="probable integral-membrane protein"
/protein_id="CAD72911.1"
/db_xref="GI:32443469"
/translation="MFFHTVLVSATLLCSLVAGFTFAEDVVAMPGIGRLNDRSFRFAF QVIDGVINQDNQPFVFWGSAIAIVAAVIGFSQIDGVDRTLLIVAAVYLLGVQLP TVRINPLNNKIQAIDVEQKDDQLLAARSEFAQWNRNVRVTTVVSITIALSLHVL LRI"
1930..2370
  /locus_tag="RB2940"
  1930..2370
  /locus_tag="RB2940"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="CAD72912.1"
  /db_xref="GI:32443470"
  /translation="MTAIAVSRLLIHLSTTGMGCLGNVACGHRDVAEVLIGYRGES EGDCEKIDPHEDGHSKRLVAAVAGSLVQHYGKTSYWELICETEPKGFHVDAIFCVSILV PLSHAWVRSGDGRSTSCNVTSPICVRRITATRELLSHETLRPHQ"
  complement(2321..2797)
  /locus_tag="RB2941"
  complement(2321..2797)
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="CAD72913.1"
  /db_xref="GI:32443471"
  /translation="MSIVSLVTAISIGAPPQISDTENVLEREDAGIQIQGRDPEAGT GLGCAFLRVPSVDVYVSLPRGFGNDARHLRLRRPPLHIQVRKGPVTEAREHELLES NVPTGAMLLYDIRLDDGNLKKRRSPROGFRFADEVDNGDIIDEDDGEIRDSANGG"
  2775..3035
  /locus_tag="RB2943"
  2775..3035
  /locus_tag="RB2943"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="CAD72914.1"
  /db_xref="GI:32443472"
  /translation="MTSHTIDIQSFQSAVHSGNIIPDRGPEVEFFQPHGKPCELSSDG CPPUCYQIGSKAKKTLTARTQFAFGQKSGSVHLLRAFRN"
  3042..3182
  /locus_tag="RB2945"
  3042..3182
  /locus_tag="RB2945"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="CAD72915.1"
  /db_xref="GI:32443473"
  /translation="MASQTVTGEQGVLRVTSCTGCKSCRRLLAEGNAKILSIKQV GS"
  3188..3334
  /locus_tag="RB2946"
  3188..3334
  /locus_tag="RB2946"
  /codon_start=1
  /transl_table=11
  /product="hypothetical protein"
  /protein_id="CAD72916.1"
  /db_xref="GI:32443474"
  /translation="MRGFAANRAAASIAVSIQQSRPNYALRSRECLGRABQORLSS KCAP"
  complement(3403..3849)
  /locus_tag="RB2948"
  complement(3403..3849)
  /locus_tag="RB2948"

```

```
/codon_start=1
/transl_table=11

Query Match      7.4%; Score 105.6; DB 1; Length 321950;
Best Local Similarity 50.9%; Pred. No. 8e-16;
Matches 330; Conservative 0; Mismatches 309; Indels 9; Gaps 3;

QY      6  CTTCCCAAGAGAACTTCAGGCTTTGTTGGTCAACCAACATCACGCCAATGACTGAGAA 65
Db      205761 CATGGCAGGCGCAAACTATCGGCGCTGATCGCGCGCAACCTACACCCCGATGAAGCGTG 205702

QY      66  TGGAGAAATCAACTTTCAGTAAATGGTCAATATGGATTATCTTGTGAAGAAACAGGG 125
Db      205701 CGCGCATCTCAACCT---CGAATGTTCTCCCGCCATGGTGGAGAGTTGCTACTTGACGG 205645

QY      126  AGTGAAGAACATTTTGTGAATGGCAACACAGGAGAGGCGCTGTCTCTGAGCGTCTCAGA 185
Db      205644  AGTCTCAGGACTGTAGTCTCGCGCGACACCGCGGAGGATGTGCTCACCACGCGAGA 205585

QY      186  GCSTCCGCCAGTTGCAAGAGTGGGTGACAAAGGAAGACAGCTGGATCAGGTGAT 245
Db      205584  ACCCCAGCTGGTCGCTCGCTTTCGTCA---ATGCGACGCGCGCGAGTTCCTCGTGTAT 205528

QY      246  AATTCAGTAGGAGCACTGAGCTTGAGGAGTCAAGGAACTGGCCCAACATCGACGAGA 305
Db      205527  CGTTCAGTTGGGCACAAACAGTTTGGGTGAGGACACAGAACTGGCCCAACATGCCCGAGCA 205469

QY      306  AATAGAGCTGATGGCATCGCTGTCATGTGACCGCTTCTTCTCAAGCCATGGACCCAAAGA 365
Db      205467  GATCGGTGCCAGTGCAGATTTCGGGACCTGCCCATCGTATTTCAAAGTTGCCAGCGTGCA 205408

QY      366  TATCCTGATTAAATTCCTAAAGGAAGTGGTGTGCGCGCCCTGCCCCCTGCCATTTTATTA 425
Db      205407  GTCGCTCACCTCTTTGCATGCAAGAACTAGCGGCGCGCCACCGGAGACGCGCTTTACTA 205348

QY      426  CTATCACATTCCTGCCCTTGACAGGGGTAAAGATTGCTGCTGAGGAGTTGTTGGATGGGAT 485
Db      205347  TTACCAACATCCCGGTGCTAACGGGTTCCAGCANTCGACATGGTGGAAATCTTGACTCAGC 205288

QY      486  TCTGGATAAGATCCCACTTCCAAAGGCTTGAATTCAGTGATACAGATCTCTTAGACTT 545
Db      205287  GGCAGAGCCATCCCGACTTTGGTGGACTGAAATACACCGATACGATGTTGTTGAGTT 205228

QY      546  CGGGCAATGTTGATCAGAAATCCCAAGCAACAGTTTGCTTTCTTTTGGGGTGGATGA 605
Db      205227  TCAACGGTGCTCGAATGTGMAATCGGA---AGTTTGAAGTCTGTTGGGGCTGCGACGA 205171

QY      606  GCAACTGTTGAGTGTCTGTGTGATGGAGCAACTGGAGCAGTGGCCAG 653
Db      205170  AATGTTATTGGGTGCGACCGCGAGCGGTGCCGTGCCGCGATTGGGAAG 205123
```

Search completed: May 11, 2004, 17:15:08
Cdb time : 5814.58 secs

GenCore version: 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 616.115 Seconds
(without alignments)
9853.151 Million cell updates/sec

Title: US-09-930-440B-1

Perfect score: 1429

Sequence: 1 atggccttcacaaagaagaa.....cttttgatcatttogatg 1429

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 3%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1429	100.0	1429	3	AA50567 Human ald
2	1429	100.0	1429	3	AA37764 Human gly
3	1429	100.0	1429	4	AH28456 Nucleotid
4	1429	100.0	1429	7	AAL53991 DNA encod
5	1332.4	93.2	2536	2	Aaz22092 cDNA enco
6	1332.4	93.2	2536	3	Aaz298699 Human N-a
7	1090.8	76.3	2441	6	ABN86096 Polyadeny
8	668.2	46.8	1171	4	AA193000 Human pol
9	652.4	45.7	1483	5	ABAC4275 Human di
10	433	30.3	444	5	AH52191 Human APP
11	352	24.6	753	4	AH63371 Human imm
12	128.2	9.0	472	8	ACH21667 Human adu
13	124	8.7	628	7	ABZ13274 Group III
14	95.4	6.7	14904	4	AAK75386 Human imm
15	77.2	5.4	618	5	ABA09731 Human bon
16	69.8	4.9	4646	6	AK68979 Human imm
17	53.4	3.7	912	6	ABN6507 Streptoco
18	53	3.7	29255	4	AA59516 Propionib
19	53	3.7	29255	7	ACF64445 Propionib
20	48.4	3.4	8056	7	ABZ10246 Haematopo
21	46.8	3.2	2000	7	ADA71938 Rice gene
22	45.4	3.2	6117	6	ABL33025 Human imm
23	44.8	3.1	5771	6	ABL33951 Human imm

24	44.4	3.1	6109	6	ABL32326	Ab132326 Human imm
25	44.4	3.1	6109	6	AA561077	AA561077 Human gen
26	44.2	3.1	1215	2	AAT58339	Aat58339 Tobacco e
27	44.2	3.1	4253	2	AAT58361	Aat58361 Tobacco e
28	44.2	3.1	19380	6	AA561426	AA561426 Human gen
29	44	3.1	8170	6	ABK28257	ABK28257 DNA trans
30	44	3.1	13574	6	ABL33317	ABL33317 Human imm
31	43.8	3.1	2000	7	ADA71938	Ada71938 Rice gene
32	43.4	3.0	915	6	ABN66506	ABN66506 Streptoco
33	43.4	3.0	110000	6	ABN71527	ABN71527 Streptoco
34	43.2	3.0	915	7	ABX07056	ABx07056 S. pneumo
35	43.2	3.0	930	7	ABZ42256	ABz42256 Streptoco
36	43.2	3.0	6361	6	ABL33141	ABL33141 Human imm
37	43.2	3.0	9897	2	AAV52143	AAV52143 Streptoco
38	43.2	3.0	17848	4	AA545323	AA545323 Chemical
39	43.2	3.0	17848	6	ABK39976	ABk39976 Human che
40	43.2	3.0	17848	6	ABK28164	ABk28164 DNA trans
41	43.2	3.0	110000	7	AB556454	Continuation (12 o
42	42.2	3.0	1075	2	AAx84335	AAx84335 Stealth v
43	42.2	3.0	8991	6	ABK31393	ABk31393 Signal tr
44	42.2	3.0	8991	6	ABL70340	ABl70340 Chemical
45	42.2	3.0	8991	6	AA561295	AA561295 Human gen

ALIGNMENTS

RESULT 1
AA50567
ID AA50567 standard; cDNA; 1429 BP.

AC AA50567;

DT 19-DEC-2000 (first entry)

DE Human aldolase cDNA.

KW Aldolase; human; sialylation; glycoprotein; plasminogen; transferrin;
KW thyrotropin; Na+,K+-ATPase; ss.

OS Homo sapiens.

EH Key Location/Qualifiers

FT CDS 1..693

FT /*tag= a

FN WO200052135-A2.

PD 08-SEP-2000.

PF 01-MAR-2000; 2000WO-US005313.

PR 02-MAR-1999; 99US-0122582P.

PR 08-DEC-1999; 99US-0169624P.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (UYJO-) UNIV JOHNS HOPKINS.

PA (UYWY-) UNIV WYOMING.

PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;

PI WPI; 2000-572178/53.

PI P-PSDB; AA56099.

PT Recombinant production of sialylated glycoproteins using cells in which

PT the expression of enzymes, e.g. sialic acid synthetase, involved in the

PT sialylation reaction has been altered.

XX Disclosure; Page 98-99; 144pp; English.

XX The present sequence is that of human cDNA encoding aldolase (see

XX AA56099). The invention provides methods and recombinantly engineered

XX cells for producing glycoproteins having sialylated oligosaccharides. The

DB	1321	TTTATACCAACTGTAACAGTCTTCAACTGTATACAACTCAAAATACACAGCTCATTT	1380
QY	1381	GGCTGCTCAGTCTAACTCTAGAAATGAGTCTTTTGAATTCATTTGCGATG	1429
DB	1391	GGCTGCTCAGTCTAACTCTAGAAATGAGTCTTTTGAATTCATTTGCGATG	1429
RESULT 3			
AAH28456			
ID	AAH28456	standard; DNA; 1429 BP.	
AC	AAH28456;		
XX	17-SEP-2001	(first entry)	
DE	Nucleotide sequence of a human aldolase.		
KW	Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;		
KW	cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;		
KW	sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;		
KW	vaccine; ss.		
XX	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	CDS	1..693	
FT		/*tag= a	
FT		/product= "aldolase"	
XX	W0200142492-AL.		
PD	14-JUN-2001.		
XX	07-DEC-2000;	2000WO-US033136.	
XX	09-DEC-1999;	99US-0169839P.	
XX	(HUMA-) HUMAN GENOME SCI INC.		
EA	(UYJO) UNIV JOHNS HOPKINS.		
EA	(UYTEM) UNIV TEMPLE.		
EA	(UYWY-) UNIV WYOMING.		
XX	Berenbaugh XJ, Lawrence S, Lee YC, Coleman TA, Palter K;		
PI	Jarvis D;		
XX	WPI; 2001-44-575/47.		
DR	P-PSDB; AAB84681.		
XX	Cells producing cytidine monophosphate-sialic acid and sialylated		
PT	glycoprotein above endogenous levels for production of vaccines and		
PT	therapeutics.		
XX	Disclosure; Fig 27; 182pp; English.		
XX	The specification describes a method for manipulating carbohydrate		
CC	processing pathways in cells of interest. The methods are used to		
CC	manipulate multiple pathways involved with the sialylation reaction by		
CC	using recombinant DNA technology and substrate feeding approaches to		
CC	enable the production of sialylated glycoproteins in the cells. The		
CC	sialylation process involves the post-translational addition of the donor		
CC	substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific		
CC	acceptor carbohydrate. The cells express at least one enzyme, selected		
CC	from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,		
CC	CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The		
CC	cells are useful for producing complex sialylated glycoproteins in cells		
CC	of interest, especially insect cells. Glycoproteins containing sialylated		
CC	oligosaccharides are useful as vaccines, therapeutics and diagnostic		
CC	tools. Cells producing complex sialylated glycoproteins are useful for		
CC	enhancing the value of heterologous expression systems and increasing the		
CC	application of heterologous cell expression products as vaccines,		
CC	therapeutics and diagnostic tools as well as increasing the variety of		
CC	heterologous proteins that can be produced and lowering biotechnology		
CC	production costs. The present sequence encodes a human aldolase, which is		

CC	used in the method of the invention		
XX	Sequence	1429 BP; 400 A; 284 C; 324 G; 421 T; 0 U; 0 Other;	
SQ	Query Match	100.0%; Score 1429; DB 4; Length 1429;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 1429; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	ATGGCCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGGCAACCATACGCCATGACT	60
DB	1	ATGGCCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGGCAACCATACGCCATGACT	60
QY	61	GGAATGGAGAAATCAACTTTTCAGTAAATGGTTCAGTATGTGATTTATCTTGTGAAGAA	120
DB	61	GGAATGGAGAAATCAACTTTTCAGTAAATGGTTCAGTATGTGATTTATCTTGTGAAGAA	120
QY	121	CAGGAGTCAAGACATTTTGTGAATGGCAACAACAGGAGAGGGCTGTCCCTGAGCGTC	180
DB	121	CAGGAGTCAAGACATTTTGTGAATGGCAACAACAGGAGAGGGCTGTCCCTGAGCGTC	180
QY	181	TCAGAGCTCGCCAGGTTGCAGAGGAGTGGGTGACAAAAGGGAAGACAAAGCTGGATCAG	240
DB	181	TCAGAGCTCGCCAGGTTGCAGAGGAGTGGGTGACAAAAGGGAAGACAAAGCTGGATCAG	240
QY	241	GTGATAATTCACGTAGGAGCACTGAGCTTGAAGGAGTCAACAGAACTGCCCAATGCA	300
DB	241	GTGATAATTCACGTAGGAGCACTGAGCTTGAAGGAGTCAACAGAACTGCCCAATGCA	300
QY	301	GCAGAAATAGGAGCTGATGGCATCGCTGCANTGCAACCGTTCTTCTCAAGCCATGGACC	360
DB	301	GCAGAAATAGGAGCTGATGGCATCGCTGCANTGCAACCGTTCTTCTCAAGCCATGGACC	360
QY	361	AAAGATATCTGATTAATTTCTAAAGGAGTGGCTGTGCGCCCGCTCCCTGCCATTT	420
DB	361	AAAGATATCTGATTAATTTCTAAAGGAGTGGCTGTGCGCCCGCTCCCTGCCATTT	420
QY	421	TATTACTATCACTTCTGCTTGACAGGGTAAAGATTCTGTCTGAGGAGTTGTTGGAT	480
DB	421	TATTACTATCACTTCTGCTTGACAGGGTAAAGATTCTGTCTGAGGAGTTGTTGGAT	480
QY	481	GGGATTTCTGGATPAAGATCCCACTTCCAAGGGCTGAAATTCAGTGATACAGATCTCTTA	540
DB	481	GGGATTTCTGGATPAAGATCCCACTTCCAAGGGCTGAAATTCAGTGATACAGATCTCTTA	540
QY	541	GACTTCGGGCAATGTGTTGATCGAATGCCAGACACAGTTTCTTCTTTTGGGGTG	600
DB	541	GACTTCGGGCAATGTGTTGATCGAATGCCAGACACAGTTTCTTCTTTTGGGGTG	600
QY	601	GATGAGCAACTGTTCAGTGTCTCTGGTATGGGAGCAACTGGAGCAGTGGSCAGTTTGT	660
DB	601	GATGAGCAACTGTTCAGTGTCTCTGGTATGGGAGCAACTGGAGCAGTGGSCAGTTTGT	660
QY	661	TCCAGAGATTTATCAACTTTGTCTCAACTAGGTTTGGAGTGTCAAGACCA	720
DB	661	TCCAGAGATTTATCAACTTTGTCTCAACTAGGTTTGGAGTGTCAAGACCA	720
QY	721	TCATCACTCTGCTCTCTGGATTCCTCAATGGGCGCCACCGGCTTCCACTGCAGAAAGCT	780
DB	721	TCATCACTCTGCTCTCTGGATTCCTCAATGGGCGCCACCGGCTTCCACTGCAGAAAGCT	780
QY	781	CCAGGAGTTTACTGATAGTGTGAAGCTAAACTGAAGAGCCTGGATTTCTTTCTTCA	840
DB	781	CCAGGAGTTTACTGATAGTGTGAAGCTAAACTGAAGAGCCTGGATTTCTTTCTTCA	840
QY	841	CTGATTTAAGATGGAACTTGGAACTGGAGCTGTAGTGTGCTCTCTATCAATCAGGT	900
DB	841	CTGATTTAAGATGGAACTTGGAACTGGAGCTGTAGTGTGCTCTCTATCAATCAGGT	900
QY	901	TTCACCTTGGACATATCTTACCTTAAATAGTGCATTTTCTCAGGAAATTTAGAT	960
DB	901	TTCACCTTGGACATATCTTACCTTAAATAGTGCATTTTCTCAGGAAATTTAGAT	960
QY	961	GAACTTGAATAAATCTCTCTAGCAAAATGAAATCTCACAATAGCAATGAGGTACCTTTTG	1020

Db 961 GAACITGAATAAATCTCTAGCAAAATGAATCTCAATAAAGCATTTGAGGTACCTTTTG 1020
QY 1021 TGAGCCTTAAAGAGTCTTATTTTGTGAAGGGGCAAAABCTTAGGAGTCAAACTCTCAG 1080
Db 1021 TGAGCCTTAAAGAGTCTTATTTTGTGAAGGGGCAAAABCTTAGGAGTCAAACTCTCAG 1080
QY 1081 TCATTCATTTACAGATTTTGTGAGAGAAATTTCTGTTTATATGGATGAATGGAATC 1140
Db 1081 TCATTCATTTACAGATTTTGTGAGAGAAATTTCTGTTTATATGGATGAATGGAATC 1140
QY 1141 AAGAGAGAAATTTGAATTAATTTCCATCTGCTTTAGAGCTCTCATATCTCGGTC 1200
Db 1141 AAGAGAGAAATTTGAATTAATTTCCATCTGCTTTAGAGCTCTCATATCTCGGTC 1200
QY 1201 TCTGGTTCTTAATCTTAATTTAAAGTTGTCTAAATTTTAAACCACTATAATATGTCAT 1260
Db 1201 TCTGGTTCTTAATCTTAATTTAAAGTTGTCTAAATTTTAAACCACTATAATATGTCAT 1260
QY 1261 TTTAATAAATATTCATTTTGAATCTAGGAAACTCTGAGCTACTGCTATTTAGGAGGCAC 1320
Db 1261 TTTAATAAATATTCATTTTGAATCTAGGAAACTCTGAGCTACTGCTATTTAGGAGGCAC 1320
QY 1321 TTTAATACCAAACTGTAACATGCTCAACTGTATACAACTCAAAATACACAGCTCATTT 1380
Db 1321 TTTAATACCAAACTGTAACATGCTCAACTGTATACAACTCAAAATACACAGCTCATTT 1380
QY 1381 GGCTGCTCAGTCTAATCTAGAAATGGATGCTTTTGAATTCATTTTCGATG 1429
Db 1381 GGCTGCTCAGTCTAATCTAGAAATGGATGCTTTTGAATTCATTTTCGATG 1429

RESULT 4

AA153391
ID AAL53391 standard; DNA; 1429 BP.
XX AAL53391;
AC AAL53391;
DT 06-MAR-2003 (first entry)
DE DNA encoding a human aldolase protein.
XX Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)Nac; sialic acid;
KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
KW transporter; sialylated glycoprotein; human; aldolase; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FT I. .693
CDS /tag= a
/product= "Human aldolase protein"
US2002142386-A1.
XX 03-OCT-2002.
XX 16-AUG-2001; 2001US-00930440.
XX 02-MAR-1999; 99US-0122582P.
XX 08-DEC-1999; 99US-016924P.
XX 25-AUG-2000; 2000US-0227579P.
XX {BETE/} BETENBAUGH M J.
XX {LAWR/} LAWRENCE S.
XX {LEEV/} LEE Y C.
XX {COLE/} COLEMAN T A.
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
XX WPI; 2003-102519/09.
XX P-PSDB; AA026544.

XX Manipulating glycoprotein production in insect cell, involves enhancing
PT expression of enzymes involved in carbohydrate processing pathway such as
PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
XX Disclosure; Fig 27; 88pp; English.
XX The invention relates to a novel method for manipulating glycoprotein
CC production in an insect cell comprising enhancing expression of an
CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
CC catalysing conversion of UDP-GlcNAc to mannose (Man)Nac, sialic acid
CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
CC synthetase or CMP-SA transporter, where the expression of each enzyme is
CC enhanced to above endogenous levels. The novel method is useful for
CC manipulating glycoprotein production in an insect cell. Further methods
CC of the invention are useful for producing sialylated glycoproteins. The
CC sialylated glycoprotein produced by the above mentioned methods are
CC useful as pharmaceutical compositions, vaccines, diagnostics and
CC therapeutics. This polynucleotide sequence represents the DNA encoding a
CC human aldolase protein of the invention
XX

SQ Sequence 1429 BP; 400 A; 284 C; 324 G; 421 T; 0 U; 0 Other;
Query Match 100.0%; Score 1429; DB 7; Length 1429;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGCAACCATCACCCATCACT 60
Db 1 ATGGCTTCCCAAGAGAAACTTCAGGGTCTTGTGGCTGCAACCATCACCCATCACT 60
QY 61 GAGATGGAGAAATCAACTTTTCAGTAATTTGGTCAAGTATGTTGATTTATTTGCAAGAA 120
Db 61 GAGATGGAGAAATCAACTTTTCAGTAATTTGGTCAAGTATGTTGATTTATTTGCAAGAA 120
QY 121 CAGGAGTGAAGAACATTTTGTGAATGGCACACAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 121 CAGGAGTGAAGAACATTTTGTGAATGGCACACAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 181 TCAGAGCTGCGCCAGGTTGCAGAGAGAGTGGTGCACAAAGAGAGAGAGAGAGAGAGAG 240
Db 181 TCAGAGCTGCGCCAGGTTGCAGAGAGAGTGGTGCACAAAGAGAGAGAGAGAGAGAG 240
QY 241 GTGATTAATTCAGTAGAGACAGTGAAGAGTCAAGAGAGTCAAGAGAGTGGCCCAACATGCA 300
Db 241 GTGATTAATTCAGTAGAGACAGTGAAGAGTCAAGAGAGTCAAGAGAGTGGCCCAACATGCA 300
QY 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTTGCAAGAGAGAGAGAGAGAGAGAG 360
Db 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTTGCAAGAGAGAGAGAGAGAGAGAG 360
QY 361 AAAGATATCTGTATTAATTTCTTAAAGAGAGTGGTGCCTGCGCCCTGCGCCATTT 420
Db 361 AAAGATATCTGTATTAATTTCTTAAAGAGAGTGGTGCCTGCGCCCTGCGCCATTT 420
QY 421 TATTACTATCAGATTTCTGCTTGAAGAGTGAAGATTCGCTGAGAGAGTGTGGAT 480
Db 421 TATTACTATCAGATTTCTGCTTGAAGAGTGAAGATTCGCTGAGAGAGTGTGGAT 480
QY 481 GGGATTTCTGATTAAGATCCCACTTCCAGAGGCTGAATTCAGTGTATCAGATCTCTTA 540
Db 481 GGGATTTCTGATTAAGATCCCACTTCCAGAGGCTGAATTCAGTGTATCAGATCTCTTA 540
QY 541 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTCTTTTGGGGTG 600
Db 541 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTCTTTTGGGGTG 600
QY 601 GATGAGCAACTGTGTAGTGTCTGTGGTATGGGAGCAACTGGAGCACTGGGCACTTTTGA 660
Db 601 GATGAGCAACTGTGTAGTGTCTGTGGTATGGGAGCAACTGGAGCACTGGGCACTTTTGA 660
QY 661 TCCAGAGATTTATCAACTTTGTGCAAACTAGTGTTCAGAGTGTCTCAGACCAAGCCCA 720

```
Dd 661 TCCAGAGATTATCAACTTTGTTGTCTCAACTAGGTTTGGAGTGTCCACAGACCAAGGCCA 720
Qy 721 TCATGACTGTGCTCTGGGATCCCAATGGGCCCCACCCGGCTTCCACTSCAGAAAGCCT 780
Dd 721 TCATGACTGTGCTCTGGGATCCCAATGGGCCCCACCCGGCTTCCACTSCAGAAAGCCT 780
Qy 781 CCAGGAGTTTACTGATAGTCTGAAGCTAACTGAAGAGCCTGGATTCCTTCTTTCA 840
Dd 781 CCAGGAGTTTACTGATAGTCTGAAGCTAACTGAAGAGCCTGGATTCCTTCTTTCA 840
Qy 841 CTSATTTAAGATGSAACCTTGAAGCTGGTAGTGTAGTGCCTCTCTATCAATCAGGCT 900
Dd 841 CTSATTTAAGATGSAACCTTGAAGCTGGTAGTGTAGTGCCTCTCTATCAATCAGGCT 900
Qy 901 TTGCACCTTGAACATTAATCTACCTTAATATAGTGAATTTTCTCAGGGAATTTAGAT 960
Dd 901 TTGCACCTTGAACATTAATCTACCTTAATATAGTGAATTTTCTCAGGGAATTTAGAT 960
Qy 961 GAACCTTGAATTAACCTCTCTAGCAATGAATCTCACATTAAGCATAGGTTACCTTTTG 1020
Dd 961 GAACCTTGAATTAACCTCTCTAGCAATGAATCTCACATTAAGCATAGGTTACCTTTTG 1020
Qy 1021 TGAGCCTTAAAGGCTTATTTTGTGAGGGGCAAAACTCTAGGAGTCAACAACCTCTCAG 1080
Dd 1021 TGAGCCTTAAAGGCTTATTTTGTGAGGGGCAAAACTCTAGGAGTCAACAACCTCTCAG 1080
Qy 1081 TCATTCATTTACAGATTTTGTGAGGAATTTCTGTTATATGATGAATGAATTC 1140
Dd 1081 TCATTCATTTACAGATTTTGTGAGGAATTTCTGTTATATGATGAATGAATTC 1140
Qy 1141 AAGAGGAATTTGATTAATTCATCTGCTTGTAGGAGCTCATATCTCGGTC 1200
Dd 1141 AAGAGGAATTTGATTAATTCATCTGCTTGTAGGAGCTCATATCTCGGTC 1200
Qy 1201 TCTGTTCTTAATCTATTTTAAAGTTGCTTAATTTTAAACCATTAATATGCTCTCAT 1260
Dd 1201 TCTGTTCTTAATCTATTTTAAAGTTGCTTAATTTTAAACCATTAATATGCTCTCAT 1260
Qy 1261 TTTAATAATATCTATTTGGAATCTAGGAATCTGAGTACTGATTTAGGAGGCAC 1320
Dd 1261 TTTAATAATATCTATTTGGAATCTAGGAATCTGAGTACTGATTTAGGAGGCAC 1320
Qy 1321 TTTAATACCAAACTGTAACCTGCTCACTGTATACAACTCAAAATACACAGCTCATTT 1380
Dd 1321 TTTAATACCAAACTGTAACCTGCTCACTGTATACAACTCAAAATACACAGCTCATTT 1380
Qy 1381 GGCTGCTCAGTCTAACTAGATGGATGCTTTTGAATTCATTTTCGATG 1429
Dd 1381 GGCTGCTCAGTCTAACTAGATGGATGCTTTTGAATTCATTTTCGATG 1429
```

RESULT 5

AAZ22092

ID AAZ22092 standard; cDNA; 2536 BP.

AC AAZ22092;

XX

DT 26-NOV-1993 (first entry)

XX

DE cDNA encoding a human N-acetylneuraminase lyase (hNANL) protein.

XX

KW Human; N-acetylneuraminase lyase; NANL; N-acetylneuraminase aldolase;

KW acylneuraminase-pyruvate lyase; sialic acid; degradation;

KW energy metabolism; sialic acid metabolism; gangliosidosis; Tay-Sachs;

KW beta-hexosaminidase deficiency; beta-galactosidase deficiency; cancer;

KW autoimmune disorder; acquired immune deficiency syndrome; AIDS;

KW Addison's disease; Crohn's disease; diabetes mellitus; atherosclerosis;

KW asthma; ss.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 89..1051

/*tag= a

US5962302-A.

05-OCT-1999.

20-FEB-1998; 98US-00027013.

20-FEB-1998; 98US-00027013.

(INCY-) INCYTE PHARM INC.

Hillman JL, Corley NC, Shah P;

WPI; 1999-571273/48.

P-PSDB; AAY40417.

Nucleic acids encoding human N-acetylneuraminase lyase useful for

treating GM1 and GM2 gangliosidosis (such as Tay-Sachs/beta-

hexosaminidase deficiency and beta-galactosidase deficiency).

Claim 4; Fig 1A-F; 31pp; English.

The present sequence encodes a human N-acetylneuraminase lyase (hNANL)

protein. NANL (also called N-N-acetylneuraminase aldolase and

acylneuraminase-pyruvate lyase) is an enzyme that catalyzes the

conversion of N-acetylneuraminic acid (sialic acid) in pyruvate and N-

acetyl-D-mannosamine (and the reverse reaction). In bacteria NANL is

involved in energy metabolism as the degradation of sialic acid produces

usable carbon and energy sources. The expression of sialic acid NANL is

regulated by the level of sialic acid. NANL is found in a number of

mammalian tissues and is a key enzyme in the metabolism of sialic acid.

hNANL polypeptides and polynucleotides may be used in the diagnosis,

prevention and treatment of disorders associated with defective sialic

acid metabolism, e.g. gangliosidosis (such as Tay-Sachs/beta-

hexosaminidase deficiency and beta-galactosidase deficiency). They may

be used to study the expression and function of NANL proteins and

their role in sialic acid metabolism. The hNANL proteins may then be used

as antigens in the production of antibodies and in assays to identify

modulators of NANL expression and activity. These antagonists of NANL may

be used to down regulate NANL expression and activity which may prove

useful in the treatment of disorders associated with over expression and

activity of NANL such as cancers (e.g. leukemia, melanoma and cancers of

the lung, brain, prostate, ovary and pancreas) and autoimmune disorders

(e.g. acquired immune deficiency syndrome (AIDS), Addison's disease,

Crohn's disease, diabetes mellitus, atherosclerosis and asthma)

Sequence 2536 BP; 744 A; 506 C; 500 G; 786 T; 0 U; 0 Other;

Query Match 93.2%; Score 1332.4; DB 2; Length 2536;

Best Local Similarity 94.3%; Fred. No. 0;

Matches 1428; Conservative 0; Mismatches 1; Indels 85; Gaps 1;

Qy 1 ATGGCTTTCCCAAGAGAAACTTCAGGGTCTTGGTGCACCAACCATCAGCCCAATGACT 60

Dd 89 ATGGCTTTCCCAAGAGAAACTTCAGGGTCTTGGTGCACCAACCATCAGCCCAATGACT 148

Qy 61 GAGATGAGAAATCAACTTTTCACTAATTTGGTCAATGATGATGATTTCTGTGAAGAA 120

Dd 149 GAGATGAGAAATCAACTTTTCACTAATTTGGTCAATGATGATGATTTCTGTGAAGAA 208

Qy 121 CAGGAGTGAAGAACTTTTGAATGCAACAGAGAGAGCTCTCCCTGAGCTC 180

Dd 209 CAGGAGTGAAGAACTTTTGAATGCAACAGAGAGAGCTCTCCCTGAGCTC 268

Qy 181 TCAGAGCTCCCGAGTTCAGAGAGAGTGGGTGCAAAAGGAGGAGCAAGCTGATCAG 240

Dd 269 TCAGAGCTCCCGAGTTCAGAGAGAGTGGGTGCAAAAGGAGGAGCAAGCTGATCAG 328

Qy 241 GTGATATTCAGTAGGAGCTGAGCTTTGAAGAGTGCACAGGAAGCTGCCCAATGCA 302

Dd 329 GTGATATTCAGTAGGAGCTGAGCTTTGAAGAGTGCACAGGAAGCTGCCCAATGCA 388

XX	SQ	Sequence	2536 BP; 744 A; 506 C; 500 G; 796 T; 0 U; 0 Other;
		Query Match	93.2%; Score 1332.4; DB 3; Length 2536;
		Best Local Similarity	94.3%; Pred. No. 0;
		Matches 1428; Conservative	0; Mismatches 1; Indels 85; Gaps 1;
QY	1	ATGGCCCTTCCCAAGAGAAACTTTCAGGTCCTGTGGCTGCAACCATCAGCCCAATGACT	60
DB	89	ATGGCCCTTCCCAAGAGAAACTTTCAGGTCCTGTGGCTGCAACCATCAGCCCAATGACT	148
QY	61	GAGAAATGGAGAAATCAACTTTTCAGTAATTTGGTCAGTATCTGGATATCTTGTGAAAGAA	120
DB	149	GAGAAATGGAGAAATCAACTTTTCAGTAATTTGGTCAGTATCTGGATATCTTGTGAAAGAA	208
QY	121	CAGGGAGTGAAGAACATTTTGTGAATGGCAACAACAGAGAGAGCCCTGTCCCTGAGCGTC	180
DB	209	CAGGGAGTGAAGAACATTTTGTGAATGGCAACAACAGAGAGAGCCCTGTCCCTGAGCGTC	268
QY	181	TCAGAGCGTCCGCCAGGTTGCAGAGGAGTGGGTGACAAAAGGGAGGACCAAGCTGGATCAG	240
DB	269	TCAGAGCGTCCGCCAGGTTGCAGAGGAGTGGGTGACAAAAGGGAGGACCAAGCTGGATCAG	328
QY	241	GTGATAATTCACGTCAGGAGCACTGAGCTTGAAGGAGTCAACAGGAACTGGGCCAACATGCA	300
DB	329	GTGATAATTCACGTCAGGAGCACTGAGCTTGAAGGAGTCAACAGGAACTGGGCCAACATGCA	368
QY	301	GCAGAAATAGGAGCTGATGGCATCGCTGCTAAAGSAAATGGCTCTGCGCGCCCTCGCCCTGCAATTT	360
DB	389	GCAGAAATAGGAGCTGATGGCATCGCTGCTAAAGSAAATGGCTCTGCGCGCCCTCGCCCTGCAATTT	448
QY	361	AAAGATATCCTGATTAATTTTCTTAAAGSAAATGGCTCTGCGCGCCCTCGCCCTGCAATTT	420
DB	449	AAAGATATCCTGATTAATTTTCTTAAAGSAAATGGCTCTGCGCGCCCTCGCCCTGCAATTT	508
QY	421	TATTACTATCAATTCCTGCTTGCAGAGGGGTAAAGATTCTGTCTGAGGAGTGTGTTGGAT	480
DB	509	TATTACTATCAATTCCTGCTTGCAGAGGGGTAAAGATTCTGTCTGAGGAGTGTGTTGGAT	568
QY	481	GGGATCTTGGATAAGATCCCCACCTTCCAGGGCTGAAATTCAGTGATACAGATCTCTTA	540
DB	569	GGGATCTTGGATAAGATCCCCACCTTCCAGGGCTGAAATTCAGTGATACAGATCTCTTA	628
QY	541	GACTTCGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTTGTCTTTTGGGCTG	600
DB	629	GACTTCGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTTGTCTTTTGGGCTG	688
QY	601	GATCAGCAACTGTTGAGTGCTCTGGTGATGGAGCAACTGAGCAAGTGGG	650
DB	689	GATCAGCAACTGTTGAGTGCTCTGGTGATGGAGCAACTGAGCAAGTGGG	748
QY	651	-----	650
DB	749	AATACCTGGGAAAAAGACAAACAGATGTTGGAGGCTTTTGAACAAAGGACTTCTCT	808
QY	651	-----CAGTTTGTATCCAGAGATTATCAACTTGTGTCAAACTAGGT	695
DB	809	TTAGCCCTGAACTATCAGTTTGTATCCAGAGATTATCAACTTGTGTCAAACTAGGT	868
QY	696	TTTGGAGTGTACAGACCAAGCCCATGACTCTGGTCTCTGGGATTCCAATGGGCCCA	755
DB	869	TTTGGAGTGTACAGACCAAGCCCATGACTCTGGTCTCTGGGATTCCAATGGGCCCA	928
QY	756	CCCGGCTTCCACTGCGAAGAGCCCTCAGGGAGTTTACTGATGCTGCTGAAGCTAAACTG	815
DB	929	CCCGGCTTCCACTGCGAAGAGCCCTCAGGGAGTTTACTGATGCTGCTGAAGCTAAACTG	988
QY	816	AAGAGCTTGGATTTCCCTTCTTTCACCTGATTTAAAGGATGCAAACTGGAGCTGGTAGC	875
DB	989	AAGAGCTTGGATTTCCCTTCTTTCACCTGATTTAAAGGATGCAAACTGGAGCTGGTAGC	1048
QY	876	TAGTGCCTCTCTATCAATCAGGTTTGCACCTTGGAGCATAACTACCTTAAATAGTGC	935

CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIFO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ

Query Match 46.8%; Score 668.2; DB 4; Length 1171;
 Best Local Similarity 89.6%; Pred. No. 6.4e-181;
 Matches 788; Conservative 0; Mismatches 3; Indels 88; Gaps 3;

QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 60
 DB 192 ATGGCTTCCCAAGAGAACTTCAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 251
 QY 61 GAGATGAGAAATCAACTTTTCAGTAATTCGTCTGATGCGATTATCTTGTGAAGAA 120
 DB 252 GAGATGAGAAATCAACTTTTCAGTAATTCGTCTGATGCGATTATCTTGTGAAGAA 311
 QY 121 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 180
 DB 312 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 371
 QY 181 TCAGAGCTGCGCAAGTTCAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGATCAG 240
 DB 372 TCAGAGCTGCGCAAGTTCAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGATCAG 431
 QY 241 GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTTCACAGGAATCGGCCCAATGCA 300
 DB 432 GTGATAATTCAGTAGAGCACTGAGCTTGAAGAGTTCACAGGAATCGGCCCAATGCA 491
 QY 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTCACCGTTCCTCTCAAGCATGGACC 360
 DB 492 GCAGAAATAGAGCTGATGGCATCGCTGTCATTCACCGTTCCTCTCAAGCATGGACC 551
 QY 361 AAAGATATCTGATTAATTTCTTAAGAGAGTGGCTGTCCGCCCTCGCCCTGCCATTT 420
 DB 552 AAAGATATCTGATTAATTTCTTAAGAGAGTGGCTGTCCGCCCTCGCCCTGCCATTT 511
 QY 421 TATTACTATCATTCCTGCTTGAAGGGTAAAGATTCGTGCTGAGAGTGTGTGAT 480
 DB 612 TATTACTATCATTCCTGCTTGAAGGGTAAAGATTCGTGCTGAGAGTGTGTGAT 671
 QY 481 GGGATCTGGATTAAGATCCCACTTCCAGGGCTGAATTCAGTACATCATCTTA 540
 DB 672 GGGATCTGGATTAAGATCCCACTTCCAGGGCTGAATTCAGTACATCATCTTA 731
 QY 541 GACTTCGGGCAATGTTGATCAGATCGCCAGCAACAGTTGCTTCTTTGGGGTG 600
 DB 732 GACTTCGGGCAATGTTGATCAGATCGCCAGCAACAGTTGCTTCTTTGGGGTG 791
 QY 601 GATGAGCAACTGTTGAGTGGCTTGGTATGGAGCAACTGGAGCAAGTGGG----- 650
 DB 792 GATGAGCAACTGTTGAGTGGCTTGGTATGGAGCAACTGGAGCAAGTGGGAGTACCTAT 851
 QY 651 ----- 650
 DB 852 AACTACCTGGGAAAAAGAACCAACAGATTTGTGGAGGCTTTTGAACAAAGGACTCTC 911
 QY 651 -----CAGTTTGTATCCAGAGATTTTATCACTTTGTTGTCAAACTAGG 694
 DB 912 TTTAGCCCTGAACATCAGTTTGTATCCAGAGATTTTATCACTTTGTTGTCAAACTAGG 971
 QY 695 TTTTGGAGTGTCAAGCAAAAGCCATCATGATCTGCTCTCT-GGGATTCAAATGGGCC 753

DB 972 TTTTGGAGTGTCAAGACCAAGCCATCATGACTCTGCTCTGGGATTCATCGGCC 1031
 QY 754 CACCCCGGCTTCCAC-TGCAGAAAGCCTCCAGGGAGTTT 791
 DB 1032 CACCCCGGCTTCCACTTCAGAAAGCCTCCAGGGAGTT 1070

RESULT 9
 ABA04275
 ID ABA04275 standard; cDNA; 1483 BP.
 XX ABA04275;
 DT 04-MAR-2002 (first entry)
 DE Human dihydrodipyrindine synthetase 31 encoding cDNA SEQ ID NO:1.
 KW Human; dihydrodipyrindine synthetase 31; malignant tumour; noschaemia;
 KW HIV infection; immunological disease; inflammation; ss.
 CS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 182..1036
 FT /*tag= a
 FT /product= "Human dihydrodipyrindine synthetase 31"

XX CN1307108-A.
 XX 08-AUG-2001.
 XX 28-JAN-2000; 2000CN-00111575.
 XX 28-JAN-2000; 2000CN-00111575.
 XX (BODA-) BODAO GENE TECH CO LTD SHANGHAI.
 PI Mao Y, Xie Y;
 XX WPI; 2002-026890/04.
 DR P-PSDB; ABB04624.
 XX New polypeptide for treating malignant tumor and HIV infection, comprises
 PT the polypeptide-dihydrodi pyridine synthetase 31 and polynucleotide for
 PT coding said polypeptide.
 XX Claim 6; Page 25-26 (Disclosure); 34pp; Chinese.

XX The present sequence encodes human dihydrodipyrindine synthetase 31 (I).
 CC (I) can be used in the treatment of various diseases, such as malignant
 CC tumour, noschaemia, HIV infection, immunological diseases and
 CC inflammations. The present invention also describes the antagonist
 CC resisting (I) and its treatment effect

XX Sequence 1483 BP; 420 A; 310 C; 344 G; 409 T; 0 U; 0 Other;
 Query Match 45.7%; Score 652.4; DB 6; Length 1483;
 Best Local Similarity 99.8%; Pred. No. 2.5e-176;
 Matches 653; Conservative 0; Mismatches 1; Indels 3; Gaps 0;

QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 60
 DB 182 ATGGCTTCCCAAGAGAACTTCAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 241
 QY 61 GAGATGAGAAATCAACTTTTCAGTAATTCGTCTGATGCGATTATCTTGTGAAGAA 120
 DB 242 GAGATGAGAAATCAACTTTTCAGTAATTCGTCTGATGCGATTATCTTGTGAAGAA 301
 QY 121 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 180
 DB 302 CAGGAGTGAAGAAATTTTGTGATGCGCAACAGAGAGAGGCTGTCTCTGAGCGTC 361
 QY 181 TCAGAGCTGCGCCAGGTTGCAAGAGAGTGGGTGACAAAGGAGGAGCAAGCTGATCAG 240

Db 362 TCAGAGCGTCGCCAGGTTGTCAGAGGAGTGGTGCACAAAGGAGGACAGCTGGATCAG 421
 QY 241 GTGATATTCACGTAGAGCACTGAGCTTGAAGAGTCAACAGGAACTGGCCCAACATGCA 300
 Db 422 GTGATATTCACGTAGAGCACTGAGCTTGAAGAGTCAACAGGAACTGGCCCAACATGCA 481
 QY 301 GCAGAAATAGAGTGTATGTCATCGCTGTCATTCACCGCTTCTTCTCAAGCCATGGACC 360
 Db 482 GCAGAAATAGAGTGTATGTCATCGCTGTCATTCACCGCTTCTTCTCAAGCCATGGACC 541
 QY 361 AAGGATATCTGATTAATTTCTTAAAGGAGTGGCTGCTGCGCCCTGCGCCCTGCGCAATT 420
 Db 542 AAGGATATCTGATTAATTTCTTAAAGGAGTGGCTGCTGCGCCCTGCGCCCTGCGCAATT 601
 QY 421 TATTACTATCACATTCCTGCTTGCAGAGGGTAAAGATTGCTGCTGAGGAGTTGTTGGAT 480
 Db 602 TATTACTATCACATTCCTGCTTGCAGAGGGTAAAGATTGCTGCTGAGGAGTTGTTGGAT 661
 QY 481 GGGATTCTGGATAAGATCCCACTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTTA 540
 Db 562 GGGATTCTGGATAAGATCCCACTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTTA 721
 QY 541 GACTTCGGGCAATGTGTGATCAGAAATGCGCCCAACAGTTTTCCTTTTGGGGTG 600
 Db 722 GACTTCGGGCAATGTGTGATCAGAAATGCGCCCAACAGTTTTCCTTTTGGGGTG 781
 QY 601 GATGCACTGTGATGCTGCTGATGGAGGACCTGGAGCAGTGGGCACT 654
 Db 782 GATGCACTGTGATGCTGCTGATGGAGGACCTGGAGCAGTGGGCACT 835

RESULT 10

ID AAH52191 standard; cDNA; 444 BP.
 XX
 AC AAH52191;
 DT 10-SEP-2001 (first entry)
 XX
 DE Human APP protein encoding cDNA sequence SEQ ID NO:197.
 XX
 KW Human; secreted protein; secretion; bacterial cell; fungal cell;
 KW eukaryotic cell; fusion protein; maltose binding protein;
 KW immunoglobulin constant region; polyhistidine tag; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200129221-A2.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000WO-US029052.
 XX
 PR 20-OCT-1999; 99US-0160712P.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Conklin DC, Yee DF;
 XX
 WP1; 2001-300340/31.
 DR P-PSDB; AAG81340.
 XX

Isolated polypeptide for directing secretion of proteins of interest from a host cell including, e.g. bacteria, includes contiguous amino acid residues of polypeptide with specified amino acids.

Claim 9; Page 347-348; 617pp; English.

AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242 to AAG81453. The secreted proteins can be used for directing the secretion of proteins of interest from a host cell including bacteria, fungal cells, and cultured higher eukaryotic cells. The present invention

CC also describes fusion proteins, where a secreted protein of the invention is operably linked via a peptide bond or peptide linker to a second protein selected from the group consisting of maltose binding protein, an immunoglobulin constant region, a polyhistidine tag and a peptide given in AAG81453

XX Sequence 444 BP; 126 A; 97 C; 117 G; 104 T; 0 U; 0 Other;

Query Match 30.3%; Score 433; DB 5; Length 444;
 Best Local Similarity 99.8%; Pred. No. 1.3e-113;
 Matches 444; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGCGCTTCCCAAGAGAAACATTCAGGGTCTTGGCTGCAACATCAAGCCCAATGACT 60
 Db 1 ATGCGCTTCCCAAGAGAAACATTCAGGGTCTTGGCTGCAACATCAAGCCCAATGACT 60
 QY 61 GAGAAATGGAGAAATCAACTTTTCAGTAAATGGTCAGTATGCGATTATCTTGTGAAGAA 120
 Db 61 GAGAAATGGAGAAATCAACTTTTCAGTAAATGGTCAGTATGCGATTATCTTGTGAAGAA 120
 QY 121 CAGGGAGTGAAGAAACATTTTGTGAATGTCACACAGGAGAGCCCTGTCCCTGAGGCTC 180
 Db 121 CAGGGAGTGAAGAAACATTTTGTGAATGTCACACAGGAGAGCCCTGTCCCTGAGGCTC 180
 QY 181 TCAGAGCTGCGCAGGTTSCAGAGGAGTGGTGTACAAAGGAGGACAGCTGGATCAG 240
 Db 181 TCAGAGCTGCGCAGGTTSCAGAGGAGTGGTGTACAAAGGAGGACAGCTGGATCAG 240
 QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAACTGGCCCAACATGCA 300
 Db 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAACTGGCCCAACATGCA 300
 QY 301 GCAGAAATAGAGCTGATGGCATCGCTGCTCATTTGCAACCGTTTCTTCTCAAGCCATGGACC 360
 Db 301 GCAGAAATAGAGCTGATGGCATCGCTGCTCATTTGCAACCGTTTCTTCTCAAGCCATGGACC 360
 QY 361 AAAGATATCTGATTAATTTCTTAAAGGAGTGGCTGCTGCGCCCTGCGCCCTGCGCAATT 420
 Db 361 AAAGATATCTGATTAATTTCTTAAAGGAGTGGCTGCTGCGCCCTGCGCCCTGCGCAATT 419
 QY 421 TATTACTATCACATTCCTGCTTGA 445
 Db 420 TATTACTATCACATTCCTGCTTGA 444

RESULT 11

ID AAK63371 standard; cDNA; 753 BP.
 XX
 AC AAK63371;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:8431.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US001354.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 15-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0285515P.
PR 07-JUN-2000; 2000US-0239467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229539P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232082P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0243207P.
PR 17-NOV-2000; 2000US-0243208P.
PR 17-NOV-2000; 2000US-0243209P.
PR 17-NOV-2000; 2000US-0243210P.
PR 17-NOV-2000; 2000US-0243211P.
PR 17-NOV-2000; 2000US-0243212P.
PR 17-NOV-2000; 2000US-0243213P.
PR 17-NOV-2000; 2000US-0243214P.
PR 17-NOV-2000; 2000US-0243215P.
PR 17-NOV-2000; 2000US-0243216P.
PR 17-NOV-2000; 2000US-0243217P.
PR 17-NOV-2000; 2000US-0243218P.
PR 17-NOV-2000; 2000US-0243244P.
PR 17-NOV-2000; 2000US-0243245P.
PR 17-NOV-2000; 2000US-0243264P.
PR 17-NOV-2000; 2000US-0243265P.
PR 17-NOV-2000; 2000US-0243297P.
PR 17-NOV-2000; 2000US-0243299P.
PR 17-NOV-2000; 2000US-0243300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR (HUMA-) HUMAN GENOME SCI INC.
PR Rosen CA, Barash SC, Ruben SM;
PR WFI; 2001-483426/52.
PR P-PSDB; AAM90590.
PR Nucleic acids encoding human immune/hematopoietic antigen polypeptides;
PR useful for preventing, diagnosing and/or treating cancers and metastasis.
PR Claim 1; SEQ ID NO 8431; 3071pp + Sequence Listing; English.
PR AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
PR amino acid sequences given in AAM91921. (I) have cytostatic
PR activity, and can be used in gene therapy and vaccine production. (I)
PR proteins and polynucleotides may be used in the prevention, diagnosis and

CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK54703
 CC to ASK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82159
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 753 BP; 195 A; 175 C; 165 G; 210 T; 0 U; 8 Other;

Query Match 24.6%; Score 352; DB 4; Length 753;
 Best Local Similarity 100.0%; Pred. No. 3e-90;

Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
WPI; 2003-615964/59.

New polynucleotide sequences obtained from various cDNA libraries, useful as hybridization probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA.

Claim 1; SEQ ID NO 8879; 44pp; English.

The invention relates to an isolated polynucleotide comprising any one of 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was determined by the technique of SBH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences are useful in diagnostics as expressed sequence tags (EST) for identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity, or in identifying mutations responsible for genetic disorders and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=200300073623

Sequence 472 BP; 134 A; 114 C; 89 G; 131 T; 0 U; 4 Other;

Query Match 9.0%; Score 128.2; DB 8; Length 472;
Best Local Similarity 97.7%; Pred. No. 4.2e-26;
Matches 130; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1297 GAGCTACTGATTAGGACAGGCATTTTAATACCAAACTGTACATCTTCAACTGTATAC 1356
61 GAATTCCTGCAATTAGGACAGGCATTTTAATACCAAACTGTACATCTTCAACTGTATAC 120

1357 AACTCAAAATACACGAGCTCATTTGGCTGCTCAGTCTAACTCTAGAAATGATGCTTTTGA 1416
121 AACTCAAAATACACGAGCTCATTTGGCTGCTCAGTCTAACTCTAGAAATGATGCTTTTGA 180

1417 ATTCATTTCGATG 1429
181 ATTCATTTCGATG 193

RESULT 13
ABZ19274/c
ID ABZ19274 standard; cDNA; 628 BP.
AC ABZ19274;
XX
XX
23-JAN-2003 (first entry)
XX
DE
DE
XX
XX
Group III cDNA cancer related clone SEQ ID NO:1700.
XX
Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.
XX Homo sapiens.
XX
XX WO200278516-A2.
PN
XX
10-OCT-2002.
PD
XX
28-MAR-2002; 2002HO-US010421.
XX
30-MAR-2001; 2001US-0280255P.
PR
28-AUG-2001; 2001US-0315563P.
PR

```
PR 09-JAN-2002; 2002US-0347313P.
XX (CORI-) CORIXA CORP.
PA Wang T, Wang S, Bangur CS, Gaiger A;
PI WPI; 2003-058387/05.
XX
DR New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and in
PT virology, immunology, microbiology, molecular biology and recombinant DNA
PT techniques.
PT
XX Claim 1; SEQ ID NO 1700; 207pp; English.
XX
XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques. N.B. The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 628 BP; 191 A; 123 C; 143 G; 171 T; 0 U; 0 Other;
Query Match 8.7%; Score 124; DB 7; Length 628;
BestLocal Similarity 100.0%; Pred.No. 7.9e-25;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 531 AGATCTCTTAGACTTCGGCAATGTTGATCAGAAATGCCAGCAACAGTTTCCT 590
Db 628 AGATCTCTTAGACTTCGGCAATGTTGATCAGAAATGCCAGCAACAGTTTCCT 569
Cy 591 TTTTGGGGTGGATGAGCAACTGTTGAGTCTCTGATGGAGCAACTGGAGCAGTGG 650
Db 568 TTTTGGGGTGGATGAGCAACTGTTGAGTCTCTGATGGAGCAACTGGAGCAGTGG 509
Cy 651 CAGT 654
Db 508 CAGT 505
RESULT 14
AAK75386
XX ID AAK75386 standard; DNA; 14504 BP.
XX AC AAK75386;
XX
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30198.
XX DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 27-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226273P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
```

PR 20-OCT-2000; 2000US-0240360P.
 PR 20-OCT-2000; 2000US-0241122P.
 PR 20-OCT-2000; 2000US-02411785P.
 PR 20-OCT-2000; 2000US-02411786P.
 PR 20-OCT-2000; 2000US-02411787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 XX Disclosure; SEQ ID NO 30198; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (II)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and

CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK37694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
 CC represent sequences used in the exemplification of the present invention
 XX
 XX SQ Sequence 14504 BP; 4313 A; 3094 C; 3219 G; 4278 T; 0 U; 0 Other;
 Query Match 6.7%; Score 95.4; DB 4; Length 14904;
 Best Local Similarity 99.0%; Pred. No. 6.7e-16;
 Matches 96; Conservative 0; Mismatches 1; Indels 0; Gaps 3;
 QY 363 AGATATCCTGATTAATTTCTTAAAGGAAGTGGCTGCGCGCCCTGCGCATTTTA 422
 |||||
 DB 13017 AGATATCCTGATTAATTTCTTAAAGGAAGTGGCTGCGCGCCCTGCGCATTTTA 13076
 QY 423 TTACTATCATTTCTGCTTGTGACAGGGTAAGATT 459
 |||||
 DB 13077 TTACTATCATTTCTGCTTGTGACAGGGTAAGATT 13113
 RESULT 15
 ABA09731
 ID ABA09731 standard; DNA; 618 BP.
 XX
 XX ABA09731;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Human bone marrow expressed oligonucleotide SEQ ID NO: 240.
 XX
 KW Human; bone marrow; cytostatic; antirheumatic; antiarthritic; vulnary;
 KW antinfammatory; antibacterial; immunosuppressive; vasotropic; cancer;
 KW antiparkinsonian; neuroprotective; nootropic; haemostatic; osteopathic;
 KW antitumor; fungicide; antidiabetic; antisthmatic; antiallergic;
 KW immunostimulant; analgesic; cerebroprotective; antianaemic; infection;
 KW nervous system disorder; autoimmune disorder; inflammation; allergy; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200174836-A1.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US010472.
 XX
 XX 31-MAR-2000; 2000US-00540217.
 XX
 XX 23-AUG-2000; 2000US-00649167.
 XX
 XX 30-NOV-2000; 2000US-0250583P.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Ford JE, Boyle BJ;
 XX WPI; 2001-626375/72.
 XX
 XX New bone marrow-expressed nucleic acids and polypeptides, useful for
 XX diagnosis, treatment of inflammatory, autoimmune, neurological, cancer
 XX and increasing hematopoiesis, stem cell survival and bone growth and
 XX remodeling.
 XX
 XX Claim 1; Page 299; 380pp; English.
 XX
 XX The present invention relates to bone marrow expressed polynucleotides
 XX and proteins. These sequences can be used in the treatment of

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 ; Search time 4131.87 Seconds
(without alignments)
10327.779 Million cell updates/sec

Title: US-09-930-440B-1

Perfect score: 1429

Sequence: 1 atggccttcctccaaagaagaa.....ctttgaattcatttcgatg 1429

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estit.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_p1n.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852.2	59.6	1015	9 AL525867	AL525867 AL525867
c 2	698.8	48.9	1201	9 AL577521	AL577521 AL577521
3	671.8	47.0	1201	9 AL554919	AL554919 AL554919
4	670.8	46.9	859	12 B1760786	B1760786 603043807

5	654	45.8	951	3	AL546960
6	643	45.0	1036	9	AL541544
7	637.2	44.6	1747	11	BC042003
c 8	608.2	42.6	709	14	CA423436
9	592.8	41.5	1403	11	AK088859
10	588.2	41.2	1403	11	AK002734
11	587.4	41.1	654	14	CK001323
12	537.2	37.6	753	9	AV728536
13	531.4	37.2	643	13	BU658094
14	531	37.2	893	14	CB999084
15	528	36.9	723	14	CB985894
16	522.4	36.6	625	9	AL597631
c 17	508	35.5	508	9	AI916625
c 18	507	35.5	508	9	AI521193
19	503.6	35.2	790	14	CK020463
20	503.6	35.2	1006	13	BY703240
21	500	35.0	635	12	BG426736
22	499.6	35.0	731	12	BI144011
23	498	34.8	975	13	BY747131
c 24	497.8	34.8	704	14	CD366774
25	496.2	34.7	693	12	BM941869
26	493.2	34.5	834	12	PI647729
27	488.6	34.2	691	14	CB057833
c 28	484	33.9	496	10	BF740024
29	478	33.4	511	14	CA396517
30	464.2	32.5	604	14	CD704049
c 31	463.4	32.4	465	9	AI635718
32	463.4	32.4	859	12	PI832112
33	459	32.1	890	10	BF691187
34	447.4	31.3	758	14	CB571304
35	446.4	31.2	774	14	CB231745
c 36	443	31.0	444	14	W78156
37	426.4	29.8	2412	11	BC034966
38	425.6	29.8	730	14	CB600330
39	420.8	29.4	876	13	BU199204
40	417.4	29.2	422	10	BF876452
41	412	28.8	646	13	BY739779
42	398	27.9	620	14	CD693284
43	392.6	27.5	669	12	BG297148
44	390.8	27.3	503	14	T87364
45	390.6	27.3	398	12	BM989844

ALIGNMENTS

RESULT 1
AL525867
LOCUS 1015 bp mRNA linear EST 23-MAY-2003
DEFINITION AL525867 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
ACCESSION AL525867
VERSION AL525867.2 GI:31063731
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
JOURNAL 1 (bases 1 to 1015)
COMMENT Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:12789360.
Contact: Genoscope
Genoscope - Centre National de Sequenage
BP 191 91006 EVRY cedex - France
Email: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 291.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODC013BF03QPl&cluster=291.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DC013BF03QPl.

FEATURES

source
1..1015
/location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC013YK06"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 59.6%; Score 852.2; DB 9; Length 1015;
Best Local Similarity 59.8%; Pred. No. 7.1e-232;
Matches 851; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTTCCCAAGAGAACTTCAGGCTCTTGTGGCTGGCAACATCATCGCCCAATGACT 60
DB 161 ATGGGCTTCCCAAGAGAACTTCAGGCTCTTGTGGCTGGCAACATCATCGCCCAATGACT 220
QY 61 GAGAAATGAGAAATCAACTTTTCAGTAACTTGTGCTAGTATGTGATTTATCTTGTGAAGAA 120
DB 221 GAGAAATGAGAAATCAACTTTTCAGTAACTTGTGCTAGTATGTGATTTATCTTGTGAAGAA 280
QY 121 CAGGAGTGAAGAACTTTTGTGAATGACCAACAGGAGAGGGCTGTCTCTGAGGCTC 180
DB 281 CAGGAGTGAAGAACTTTTGTGAATGACCAACAGGAGAGGGCTGTCTCTGAGGCTC 340
QY 181 TCAGAGCTGCGCCAGTGTGCAGAGAGTGGTGCACAAAGGAGGAGGAGGAGGAGGAGGAGG 240
DB 341 TCAGAGCTGCGCCAGTGTGCAGAGAGTGGTGCACAAAGGAGGAGGAGGAGGAGGAGGAGG 400
QY 241 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGAGTGCACAGGAACTGCGCCCAACATGCA 300
DB 401 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGAGTGCACAGGAACTGCGCCCAACATGCA 460
QY 301 GCAGAAATAGAGCTGATGGGATCGCTGTCAATGCAACCGTTCTTCTCAAGCCATGGACC 360
DB 461 GCAGAAATAGAGCTGATGGGATCGCTGTCAATGCAACCGTTCTTCTCAAGCCATGGACC 520
QY 361 AAAGATATCTGATTAATTTCTTAAGGAGTGGTGTGCGCCGCGCCCTGCCCTGCCATTT 420
DB 521 AAAGATATCTGATTAATTTCTTAAGGAGTGGTGTGCGCCGCGCCCTGCCCTGCCATTT 580
QY 421 TATTACTATCATCTCTGCTGTGACAGGAGTAAAGATTCGTGCTGAGGAGTGTGGAT 480
DB 581 TATTACTATCATCTCTGCTGTGACAGGAGTAAAGATTCGTGCTGAGGAGTGTGGAT 640
QY 481 GGGATTCGATAGATCCCACTTCCAGGGCTGAATTCAGTGATACATCTCTTA 540
DB 641 GGGATTCGATAGATCCCACTTCCAGGGCTGAATTCAGTGATACATCTCTTA 700
QY 541 GACTTCGCGCAATGCTGATCAGAAATCGCCAGCAACAGTTGCTTCTTTCGGGGT 600
DB 701 GACTTCGCGCAATGCTGATCAGAAATCGCCAGCAACAGTTGCTTCTTTCGGGGT 760
QY 601 GATGAGCAACTGTTGAGTGTCTTGTGATGGAGCAACTGGAGCAGTGGGAGTTTGTGA 660
DB 761 GATGAGCAACTGTTGAGTGTCTTGTGATGGAGCAACTGGAGCAGTGGGAGTTTGTGA 820
QY 661 TCCAGAGATTTATCAACTTTGTGTCAACTAGGTTTGGAGTGTCAAGAACCAAGGCA 720
DB 821 TCCAGAGATTTATCAACTTTGTGTCAACTAGGTTTGGAGTGTCAAGAACCAAGGCA 880
QY 721 TCATGACTCTGCTCTCGGATTTCCAAATGCGCCCAACCGGCTTCCACTGGAGAGGCT 780
DB 881 TCATGACTCTGCTCTCGGATTTCCAAATGCGCCCAACCGGCTTCCACTGGAGAGGCT 940
QY 781 CCAGGAGTTTACTGATAGTGTGAAGCTAAACTGAAGAGCTGTGATTTCTTTCTTCA 840

DB 941 CCAGGAGTTTACTGATAGTGTGAGCTAACTGAGAGCCCTGGATTCTTCTTCA 1600

QY 841 CTGATTAAAGGA 853

DB 1001 CTGATTAAAGRK 1013

RESULT 2

AL577521/c

LOCUS

DEFINITION

AL577521 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

clone CS0DI08YB23 3-PRIME, mRNA sequence.

AL577521

ACCESSION

AL577521

VERSION

AL577521.2

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 1201)

AUTHORS

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On Feb 16, 2001 this sequence version replaced gi:12940733.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 92006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 291.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgl-bin/cluster.cgi?seq=CS0DI087CA12NP1&cluster=291.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DI087CA12NP1.

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DI08YB23"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 48.9%; Score 696.2; DB 9; Length 1201;

Best Local Similarity 83.2%; Pred. No. 5.3e-188;

Matches 844; Conservative 16; Mismatches 68; Indels 87; Gaps 3;

QY 456 GATTCGCTGCTGAGGAGTGTGGATGGGATTCGATGAAGATCCACACCTTCCAGGGCT 515

DB 1018 GATTCGCGGATGAGGAGTGTGGATGGGATTCGATGAAGATCCACACCTTCCAGGGCT 959

QY 516 GAAATTCAGTATCAGATCTCTTAGACTTCGGGCAATGTGTGATCAGAAATCGCCAGCA 575

DB 958 GAAATTCAGTATCAGATCTCTTAGACTTCGGGCAATGTGTGATCAGAAATCGCCAGCA 899

QY 576 ACAGTTCTCTTCTTTTGGGTCGATGAGCACTGTGATGCTCTGTTGATGGGAGC 635

DB 898 ACAGTTCTCTTCTTTTGGGTCGATGAGCACTGTGATGCTCTGTTGATGGGAGC 839

QY 636 AACTGGAGCAGTGGG----- 650

DB 838 SACTGGAGCAGTGGGAGTACCTATCTACTCTGGGAAAAAAGACAAACCGGTGTGGCA 779

QY 651 -----CAGTTTGTATCCAGAGATT 670

Db 778 RSCVTTTGAACAAGAACTCTCTTTAGCCCTGAACTHACAGTTTGTATCCAGAGHT 719
 Qy 671 TATCAACTTTTGTGTAACATAGGTTTGGAGTGTCAAGAGCCAAAGCCATCATGACTCT 730
 Db 718 TATCAACTTTTGTGTAACATAGGTTTGGAGTGTCAAGAGCCAAAGCCATCATGACTCT 659
 Qy 731 GGTCTCTGGGATTCAGTGGGCGCCACCCGG-GCTTCCACTGAGAAAGCCCTCCAGGAGT 789
 Db 658 GGTCTCTGGGATTCAGTGGGCGCCACCCGGCTTCCACTGAGAAAGCCCTCCAGGAGT 599
 Qy 790 TTACTGATAGTCTGAAGCTAAAGTGAAGAGGCTGGATTCCTTTCTTTTCACTGATTTAA 849
 Db 598 TNACTGATAGTCTGAAGCTAAAGTGAAGAGGCTGGATTCCTTTCTTTTCACTGATTTAA 539
 Qy 850 AGGATGGAACTTTGTAAGGCGGAAACTCTAGTGTCTCTATCAAAATCAGGGTTTGACCTT 909
 Db 538 AGGATGGAACTTTGTAAGGCGGAAACTCTAGTGTCTCTATCAAAATCAGGGTTTGACCTT 479
 Qy 910 GAGACATAACTTACTTAAATAGTGCATTTTCTCTCAGGGAATTTAGATGAACCTGAA 969
 Db 478 GAAACATAACTTACTTAAATAGTGCATTTTCTCTCAGGGAATTTAGATGAACCTGAA 419
 Qy 970 TAAACTCTCTAGCAAAATGAAATCTCAAAATAGCAATCAGGTACCTTTTGTAGGCTTTA 1029
 Db 418 TAAACTCTCTAGCAAAATGAAATCTCAAAATAGCAATCAGGTACCTTTTGTAGGCTTTA 359
 Qy 1030 AAAAGTCTTATTTGTGAAGGGGAAACTCTAGGAGTCAAACTCAGTCATCATTCATT 1089
 Db 358 AAAAGTCTTATTTGTGAAGGGGAAACTCTAGGAGTCAAACTCAGTCATCATTCATT 299
 Qy 1090 TCACAGATTTTGTGAGAGAAATTCGTTTATATGATGAATGAATGAAGGAA 1149
 Db 298 TCACAGATTTTGTGAGAGAAATTCGTTTATATGATGAATGAATGAAGGAA 239
 Qy 1150 ATTGTAATTTGTAATTCATCTGTCTTTAGGAGCTCTCATTAICTGGTCTCTGGTTCC 1209
 Db 238 ATTGTAATTTGTAATTCATCTGTCTTTAGGAGCTCTCATTAICTGGTCTCTGGTTCC 179
 Qy 1210 TAATCTTATTTAAAGTTGTCTAATTTAAACCACTCAATATGTCCTTCAATTTAATAA 1269
 Db 178 TAATCTTATTTAAAGTTGTCTAATTTAAACCACTCAATATGTCCTTCAATTTAATAA 119
 Qy 1270 TAATCTTATTTAAAGTTGTCTAATTTAAACCACTCAATATGTCCTTCAATTTAATAA 1328
 Db 118 TAATCTTATTTAAAGTTGTCTAATTTAAACCACTCAATATGTCCTTCAATTTAATAA 59
 Qy 1329 CAAACTGTAACTGTCTCAACTGTATACAACTCAAAATACACCAGCTCATTTTGGC 1383
 Db 58 CAAACTGTAACTGTCTCAACTGTATACAACTCAAAATACACCAGCTCATTTTGGC 4

RESULT 3
 AL554919
 LOCUS
 DEFINITION AL554919 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 Clone CS0D1087B23 5-PRIME, mRNA sequence.
 ACCESSION AL554919
 VERSION AL554919.2 GI:31276729
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1201)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On Feb 15, 2001 this sequence version replaced gi:12896160.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 291.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0D1087CA12QP1&cluster=291.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0D1087CA12QP1.
 Location/Qualifiers
 1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1087B23"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="First strand cDNA was primed with a NotI-cligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES

source

ORIGIN

Query Match 47.0%; Score 671.8; DB 9; Length 1201;
 Best Local Similarity 88.9%; Pred. No. 2.8e-180;
 Matches 792; Conservative 4; Mismatches 7; Indels 88; Gaps 3;
 QY 465 TGAGGAGTTGTTGGATGGGATTCGGTAAAGATCCCACTTCCAGGGCTGAATTCAG 524
 DB 69 TGAGGAGTTGTTGGATGGGATTCGGTAAAGATCCCACTTCCAGGGCTGAATTCAG 128
 QY 535 TGATACAGATCTCTTAGACTTCGGGCAATGCTTGATCAGATCGCACCAACAGTTTC 594
 DB 129 TGATACAGATCTCTTAGACTTCGGGCAATGCTTGATCAGATCGCACCAACAGTTTC 188
 QY 585 TTTCTTTTGGGTTGGATGAGCAACTGTTGAGTCTCTGGTGGGAGCAACTGGAGC 644
 DB 189 TTTCTTTTGGGTTGGATGAGCAACTGTTGAGTCTCTGGTGGGAGCAACTGGAGC 248
 QY 645 AGTGGG----- 650
 DB 249 AGTGGGAGTACCTATACTACCTGGGAAAAAAGACAAACCAAGNTGTTGGAGCTTTTGA 308
 QY 651-----CAGTTTGTATCCAGAGATTTTATCAACT 679
 DB 309 ACAAAGAGACTCTCTTTAGCCCTGAATCAGTTTGTATCCAGAGATTTTATCAACT 368
 QY 680 TGTGTCAAACTAGGTTTGGAGTGTACAGACCAAGCCATCATCTGCTCTGG 739
 DB 369 TGTGTCAAACTAGGTTTGGAGTGTACAGACCAAGCCATCATCTGCTCTGG 428
 QY 740 GATTCCAAATGGGCCACCCCGCTTCCACTGAGAAAGCCCTCCAGGGAGTTTACTGATAG 799
 DB 429 GATTCCAAATGGGCCACCCCGCTTCCACTGAGAAAGCCCTCCAGGGAGTTTACTGATAG 488
 QY 800 TGCTGAAGCTAACTGAAGAGCTGGATTCCTTTCTTTCTACTGATTTAAAGATGGAAA 859
 DB 489 TGCTGAAGCTAACTGAAGAGCTGGATTCCTTTCTTTCTACTGATTTAAAGATGGAAA 548
 QY 860 CTGGAAGCTGGTACTAGTGCCTCTCTATCAAAATCAGGGTTTGACCTTTGAGACATAT 919
 DB 549 CTGGAAGCTGGTACTAGTGCCTCTCTATCAAAATCAGGGTTTGACCTTTGAGACATAT 608
 QY 920 CTACCTTAAATAGTGCATTTTCTCAGGGAATTTAGATGAACCTGAATAAATCTCC 979
 DB 609 CTACCTTAAATAGTGCATTTTCTCAGGGAATTTAGATGAACCTGAATAAATCTCC 668
 QY 980 TAGCAATGAATCTCAAAATAGCAATGAGTACCTTTTGTAGCCCTTAAAGTCTTA 1039
 DB 669 TAGCAATGAATCTCAAAATAGCAATGAGTACCTTTTGTAGCCCTTAAAGTCTTA 728
 QY 1040 TTTTGTGAAGGGGCAAAACTCTAGGAGTCAAACTCTCAGTCATTCATTTACAGATTT 1099
 DB 729 TTTTGTGAAGGGGCAAAACTCTAGGAGTCAAACTCTCAGTCATTCATTTACAGATTT 788

```

QY 1100 TTTTCGGAGAAATTTCTCTTTATATGGAATGAAATCAAGAGAAATGTAAATG 1159
Db 789 TTTTCGGAG-AAATTTCTCTTTATATGGAATGAAATCAAGAGAAATGTAAATG 847
QY 1160 ATTTACTCCATGCTGCTTTAGGAGCTCTCATTTATCTCGGTCTCTGGTTCCTTAAT 1219
DE 848 ATTAACTCCATGCTGCTTTAGGAGCTCTCATTTATCTCGGTCTCTGGTTCCTTAAT 907
QY 1220 TTAAGCTTCTTAATTTAAACACATATATATGCTTCTCATTTTAATAAT 1270
Db 908 TTAAGCTTCTTAATTT--AACACATATATATGCTTCTCATTTTAATAAT 956

RESULT 4
LOCUS B1760786 859 bp mRNA linear EST 25-SEP-2001
DEFINITION 603043807F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5184192 5',
mRNA sequence.
ACCESSION B1760786
VERSION B1760786.1 GI:15723364
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rstraus@nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11459 row: n column: 01
High quality sequence stop: 858.
FEATURES
Location/Qualifiers
1..855
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5184192"
/lab_host="DH10B"
/clone_lib="NIH MGC 116"
/note="Organ: pCooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site1: NotI; Site2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 45.3%; Score 670.8; DB 12; Length 859;
Best Local Similarity 97.5%; Pred. No. 4.6e-180;
Matches 745; Conservative 0; Mismatches 12; Indels 7; Gaps 6;
QY 1 ATGGCTTCCCAAGAGAAACTTCAGGCTTTGTGCTGCAACCATCAGCCCAATGACT 60
Db 99 ATGGCTTCCCAAGAGAAACTTCAGGCTTTGTGCTGCAACCATCAGCCCAATGACT 158
QY 61 GAGAAATGGAATAATCACTTTTCAGTAAATGGTCAGTATGTGATTAATCTGTGAAAGAA 120
Db 159 GAGAAATGGAATAATCACTTTTCAGTAAATGGTCAGTATGTGATTAATCTGTGAAAGAA 218

```

```

QY 121 CAGGGAGTGAAGAACATTTTGTGTAATGGCAACACAGAGAGAGGCTCTCCCTGAGGCTC 180
Db 219 CAGGGAGTGAAGAACAA-TTTTGTGAATGGCAACACAGAGAGAGGCTCTCCCTGAGGCTC 277
QY 181 TCAGAGCGTCGCCAGGCTTCAGAGGAGTGGGTGACAAAAGGGAAGGACAAAGCTGGAATCAG 240
Db 278 TCAGAGCGTCGCCAGGCTTCAGAGGAGTGGGTGACAAAAGGGAAGGACAAAGCTGGAATCAG 337
QY 241 GTGATAATTCACGTAGGAGACATGAGCTTTGAAGGAGTCAAGAACTGGCCCAACATGCA 300
Db 338 GTGATAATTCACGTAGGAGACATGAGCTTTGAAGGAGTCAAGAACTGGCCCAACATGCA 397
QY 301 GCAGAAATAGGAGCTGATGGCATCGCTGTCATTCGACCGTTCCTTCCTCAAGCCATGACC 360
Db 398 GCAGAAATAGGAGCTGATGGCATCGCTGTCATTCGACCGTTCCTTCCTCAAGCCATGACC 457
QY 361 AAAGATATCTCTGATTAATTTCCCTAAAGAGAGTGGCTGTGCTGCCGCCCTGCTGCCCATTT 420
Db 458 AAAGATATCTCTGATTAATTTCCCTAAAGAGAGTGGCTGTGCTGCCGCCCTGCTGCCCATTT 517
QY 421 TATTACTATCAATTCCTGCTTGAAGAGTGAAGATTCGCTGCTGAGGAGTGTGGAT 480
Db 518 TATTACTATCAATTCCTGCTTGAAGAGTGAAGATTCGCTGCTGAGGAGTGTGGAT 576
QY 481 GGGATTCCTGGATTAAGTCCCACTTCCAGGGGTGAATTCAGTGTGATCAGATCTCTTA 540
Db 577 GGGATTCCTGGATTAAGTCCCACTTCCAGGGGTGAATTCAGTGTGATCAGATCTCTTA 636
QY 541 GACTTCGGGCAATGTGTTGATCAGAAATCCAGCAACAGTTTCTTTCTTTTTCCTTTT-GGGGT 599
Db 637 GACTTCGGGCAATGTGTTGATCAGAAATCCAGCAACAGTTTCTTTCTTTTTCCTTTT-GGGGT 696
QY 600 GATGAGCAACTGTTCAGTGTCTGCTGTGATGGAGCAACT-GGAGAGTGGGAGTTTG 658
Db 697 GATGAGCAACTGTTCAGTGTCTGCTGTGATGGAGCAACTGGGAGCAGTGGGAGTTTG 756
QY 659 TATCCAGAGATTTATCAACTTTGTCGTAACCTAGGTTTGGAGTGTACAGACCAAGC 718
Db 757 TATCCAGAGATTTATCAACTTTGTCGTAACCTAGGTTTGGAGTGTACAGACCAAGC 815
QY 719 CATCATGACTCTGG--TCTTGGGATTCCAATGGGCCACCCG 760
Db 816 CATCATGACTCTGGGCTCTCGGGATTCCAATGGGCCACCCG 859

RESULT 5
LOCUS AL546960
DEFINITION AL546960 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1032XJ10 5-PRIME, mRNA sequence.
ACCESSION AL546960
VERSION AL546960.2 GI:31268793
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 951)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:12880585.
Contact: Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 291.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1032XJ10&cluster=291.f. Contact :
Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

```

```

FEATURES
source
Faraday Avenue Genoscope sequence ID : CS0DI032DE05QPL.
Location/Qualifiers
1..951
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI032YU10"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 45.8%; Score 654; DB 9; Length 951;
Best Local Similarity 100.0%; Pred. No. 3.1e-175;
Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCTCCCAAGAGAACTTCCAGGCTCTTGGCTGCAACCATCAAGCCATGACT 60
DB 257 ATGGCCTCCCAAGAGAACTTCCAGGCTCTTGGCTGCAACCATCAAGCCATGACT 316
QY 61 GAGATGGAGAAATCACTTTTCAGTAATTCGTAGTATGGATTATCTTGTGAAGAA 120
DB 317 GAGATGGAGAAATCACTTTTCAGTAATTCGTAGTATGGATTATCTTGTGAAGAA 376
QY 121 CAGGAGTGAAGACATTTTGTGAATGGCAACAGGAGAGGCTTCCTCGAGCGTC 180
DB 377 CAGGAGTGAAGACATTTTGTGAATGGCAACAGGAGAGGCTTCCTCGAGCGTC 436
QY 181 TCAGAGCTGCCAGGTGCAAGAGAGTGGTGTCAAAAGGAGAGCAAGCTGATGAC 240
DB 437 TCAGAGCTGCCAGGTGCAAGAGAGTGGTGTCAAAAGGAGAGCAAGCTGATGAC 496
QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAATGGCCCAATGCA 300
DB 497 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGTCAACAGGAATGGCCCAATGCA 556
QY 301 GCAGAAATAGAGCTGATGGCATCGCTGTCATTGACCGCTTCTTCCCTCAAGCCATGAGCC 360
DB 557 GCAGAAATAGAGCTGATGGCATCGCTGTCATTGACCGCTTCTTCCCTCAAGCCATGAGCC 616
QY 361 AAGAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCCTGCAATTT 420
DB 617 AAGAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCCTGCAATTT 676
QY 421 TATTACTATCAATTCCTGCTTGCAGAGGCTAAAGATTGCTGCTGAGGAGTGTGTTGAT 480
DB 677 TATTACTATCAATTCCTGCTTGCAGAGGCTAAAGATTGCTGCTGAGGAGTGTGTTGAT 736
QY 481 GGGATTCTGATAGATCCCACTTCCAGAGGCTGAAATTCAGTGATACAGATCTCTTA 540
DB 737 GGGATTCTGATAGATCCCACTTCCAGAGGCTGAAATTCAGTGATACAGATCTCTTA 796
QY 541 GACTTCGGGCAATGTGTGATCAGATCGCCACCAACAGTTGCTTCTTTTGGGGTG 600
DB 797 GACTTCGGGCAATGTGTGATCAGATCGCCACCAACAGTTGCTTCTTTTGGGGTG 856
QY 601 GATGAGCAATGTGTGATGCTCTGGTGATGGGAGCAACTGGAGCAGTGGGCACT 654
DB 857 GATGAGCAATGTGTGATGCTCTGGTGATGGGAGCAACTGGAGCAGTGGGCACT 910

RESULT 6
AL541544 1036 bp mRNA linear EST 12-MAY-2003
LOCUS AL541544 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE003YCI9
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION AL541544
VERSION AL541544.2 GI:30545825
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1036)

AUTHORS

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On Feb 15, 2001 this sequence version replaced gi:12872718.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 291.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DE003AB10P1&cluster=291.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DE003AB10QPL.

FEATURES

Location/Qualifiers

1..1036

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DE003YCI9"

/tissue_type="PLACENTA"

/clone_lib="Homo sapiens PLACENTA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo (dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

ORIGIN

Query Match 45.0%; Score 643; DB 9; Length 1036;

Best Local Similarity 99.8%; Pred. No. 4.4e-172;

Matches 654; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGCCTTCCCAAGAGAACTTCCAGGCTCTTGGCTGCAACCATCAAGCCATGACT 60

DB 234 ATGGCCTTCCCAAGAGAACTTCCAGGCTCTTGGCTGCAACCATCAAGCCATGACT 293

QY 61 GAGATGGAGAAATCACTTTTCAGTAATTCGTAGTATGGATTATCTTGTGAAGAA 120

DB 294 GAGATGGAGAAATCACTTTTCAGTAATTCGTAGTATGGATTATCTTGTGAAGAA 353

QY 121 CAGGAGTGAAGACATTTTGTGAATGGCAACAGAGAGGCTTCTCCCTGAGCGTC 180

DB 354 CAGGAGTGAAGACATTTTGTGAATGGCAACAGAGAGGCTTCTCCCTGAGCGTC 413

QY 181 TCAGAGCTGCCAGGTTGCAGAGGAGTGGTGTGCAAAAGGAGGACAGCTGATCAG 240

DB 414 TCAGAGCTGCCAGGTTGCAGAGGAGTGGTGTGCAAAAGGAGGACAGCTGATCAG 473

QY 241 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGAGTCAAGAACTGGCCCAATGCA 300

DB 474 GTGATAATTCACGTAGGAGCACTGAGCTTGAAGAGAGTCAAGAACTGGCCCAATGCA 533

QY 301 GCAGAAATAGGAGCTGATGGCATCGCTGTCATTGCAACCGTTCTTCCCTCAAGCCATGAGCC 360

DB 534 GCAGAAATAGGAGCTGATGGCATCGCTGTCATTGCAACCGTTCTTCCCTCAAGCCATGAGCC 593

QY 361 AAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCCTGCAATTT 420

DB 594 AAGATATCTGATTAATTTCTTAAGAGAGTGGCTGCTGCCGCCCTGCTGCCCTGCAATTT 653

QY 421 TATTACTATCAATTCCTGCTTGCAGAGGCTAAAGATTGCTGCTGAGGAGTGTGTTGGA 479

DB 654 TATTACTATCAATTCCTGCTTGCAGAGGCTAAAGATTGCTGCTGAGGAGTGTGTTGGA 713

QY 480 TGGGATCTGGGATAGATGCCCAACCTTCCAGGGCTGAAATTCAGTGATACAGATCTCTT 539

```

Db      714 TGGATTCTGGATAGATCCCAACCTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTT 773
Qy      540 AGACTTCGGGCAATGCTGTGATCAGAAATCGGCAGCAACAGTTTCTTTCCTTTTGGGGT 599
Db      774 AGACTTCGGGCAATGCTGTGATCAGAAATCGGCAGCAACAGTTTCTTTCCTTTTGGGGT 833
Qy      600 GGATGAGCACTGTGTAGTGCCTCTGTGATGGAGCACTGAGAGCAGTGGGCGT 654
Db      834 GGAATGAGCACTGTGTAGTGCCTCTGTGATGGAGCACTGAGAGCAGTGGGCGT 888

RESULT 7
LOCUS   BC042003              1747 bp      mRNA      linear      HTC 02-JAN-2003
DEFINITION Homo sapiens, clone IMAGE:5311213, mRNA.
ACCESSION BC042003
VERSION   BC042003.1 GI:27469544
KEYWORDS HTC.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1. (bases 1 to 1747)
AUTHORS   Strausberg, J.
TITLES    Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNALS  Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
          2. (bases 1 to 1747)
          Strausberg, J.
          Submitted (23-DEC-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
          NIH-MGC Project URL: http://mgc.nci.nih.gov
          Contact: MGC help desk
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: Michael J. Brownstein (NECRI) & Shiraki
          Toshiyuki and Piero Carninci (RIKEN)
          cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          Center, Stanford University School of Medicine, Stanford, CA 94305
          Web site: http://www.shgc.stanford.edu
          Contact: (Dickson, Mark) mcdpaxil.stanford.edu
          Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
          R. M.
          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK plate: 77 Row: n Column: 20
          This clone has the following problem: frame shifted.

FEATURES             source
    source            1..1747
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5311213"
                     /tissue_type="Brain, hypothalamus"
                     /clone_lib="NIH MGC_96"
                     /lab_host="DH10B"
                     /note="vector: pBluescript"

ORIGIN
Query Match      44.6%; Score 637.2; DB 11; Length 1747;
Best Local Similarity 99.4%; Pred. No. 2.6e-170;
Matches 650; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy      1 ATGGCTTCCCAAGAGAACTTCAGGGTCTTGGCTGCAACCAATCAAGCCCAATGACT 60
Db      429 ATGGCTTCCCAAGAGAACTTCAGGGTCTTGGCTGCAACCAATCAAGCCCAATGACT 488
Qy      61 GAGATGGAGAAATCACTTTTCAGTAATGTTCAGTATGTGATATCTCTGCAAGAA 120
Db      489 GAGATGGAGAAATCACTTTTCAGTAATGTTCAGTATGTGATATCTCTGCAAGAA 548
Qy      121 CAGGGAGTGAAGAACATTTTGTGTAATGGCACAACAGGAGAGCCCTGCTCCCTGAGCGTC 180

```

```

Db      549 CAGGAGTGAAGAACATTTTGTGATCGCACACACAGGAGAGGCTGTCCCTGAGCGTC 608
Qy      181 TCAGAGCTCCCCAGGTTGCAGAGAGTGGGTGA CAAAGGGAAGGACAGCTGATCAG 240
Db      609 TCAGAGCTCCCCAGGTTGCAGAGAGTGGGTGA CAAAGGGAAGGACAGCTGATCAG 668
Qy      241 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGGAGTGCACAGAACTGSCCAACATGCA 300
Db      669 GTGATAATTCAGTAGGAGCACTGAGCTTGAAGGAGTGCACAGAACTGSCCAACATGCA 728
Qy      301 GCAGAAATAGGAGCTGATGGCATCGCTGTCTATTGACCCGTTCTTCTCAAGCAATGGACC 360
Db      729 GCAGAAATAGGAGCTGATGGCATCGCTGTCTATTGACCCGTTCTTCTCAAGCAATGGACC 788
Qy      361 AAAGATATCTGATTAATTTCTTAAAGGAAGTGGCTGCTGCCGCCCTGCCCCGCAATTT 420
Db      789 AAAGATATCTGATTAATTTCTTAAAGGAAGTGGCTGCTGCCGCCCTGCCCCGCAATTT 848
Qy      421 TATTACTATCATTTCTGCTTGCAGAGGGTAAAGATTCTGCTGAGGAGTTGTGGAT 480
Db      849 TATTACTATCATTTCTGCTTGCAGAGGGTAAAGATTCTGCTGAGGAGTTGTGGAT 908
Qy      481 GGGATTCTGGATAAGATCCCAACCTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTTA 540
Db      909 GGGATTCTGGATAAGATCCCAACCTTCCAAAGGGCTGAAATTCAGTGATACAGATCTCTTA 968
Qy      541 GACTTGGGCAATGTGTGATCAGATCCGCAAGCAACAGTTTGTCTTCTTCTTTGGGGTG 600
Db      969 GACTTGGGCAATGTGTGATCAGATCCGCAAGCAACAGTTTGTCTTCTTCTTTGGGGTG 1028
Qy      601 GATGAGCACTGTTCAGTGTCTCTGCTGATGGAGCAACTGGAGCACTGGGCACT 654
Db      1029 GATGAGCACTGTTCAGTGTCTCTGCTGATGGAGCAACTGGAGCACTGGGCACT 1081

RESULT 8
LOCUS   CA423436/c              709 bp      mRNA      linear      EST 07-NOV-2002
DEFINITION UI-H-FEI-bec-9-13-0-UI.s1 NCI_CGAP_FEI Homo sapiens cDNA clone
ACCESSION UI-H-FEI-bec-9-13-0-UI.3', mRNA sequence.
VERSION   CA423436
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1. (bases 1 to 709)
          NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
          Unpublished (1997)
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: James Martin
          cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
          DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: Clone distribution information can be obtained
          from Dr. M. Bento Soares, bentto-soares@uiowa.edu
          Seq primer: M13 FORWARD
          POLYA=Yes.
          Location/Qualifiers
            source      1..709
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-FEI-bec-9-13-0-UI"
                     /tissue_type="Cell lines"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_FEI"
                     /note="Organ: Chondrosarcoma; Vector: pT7T3-Pac"

```

(Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP FEI is a normalized cDNA library derived from a pool of mRNA obtained from 3 cell lines from grade II chondrosarcoma tissues. The library was constructed according to Bonaleo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCTACGGAC. The cell lines were provided by Dr James Martin from the University of Iowa.
TAG TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG LIB=UI-H-FEI
TAG_SEQ=CGCTACGGAC"

ORIGIN

Query Match 42.6%; Score 608.2; DB 14; Length 709;
Best Local Similarity 98.8%; Pred. No. 3.3e-162;
Matches 655; Conservative 0; Mismatches 3; Indels 5; Gaps 4;

Qy	341	TCCTCTCAAGCCATGACCAAGATATCCGATTAATTTCTTAAGGAAGTGCCTGCT	400
Db	670	TCCTCTCAAGCCATG--ACCAAGATATCC--GA--TAATTTCTTAAGGAAGTGG--TGCTG	615
Qy	401	CGGCCCCCTGGCTCCATTTTATTACTATCATCTCCCTGCTTGAC--AGGGGTAAAGATT	459
Db	614	CGGCCCCCTGGCTCCATTTTATTACTATCATCTCCCTGCTTGAC--AGGGGTAAAGATT	555
Qy	460	CGTCTCAGGAGTTGTTGGATGGATTTCTGGATAAGATCCCAACCTTCCAAAGGGCTGAAA	519
Db	554	CGTCTCAGGAGTTGTTGGATGGATTTCTGGATAAGATCCCAACCTTCCAAAGGGCTGAAA	495
Qy	520	TTCAAGTATACAGATCTTTAGACTTCGGGAAATGTGTGTATCAGATCCGACGACACAG	579
Db	494	TTCAAGTATACAGATCTTTAGACTTCGGGAAATGTGTGTATCAGATCCGACGACACAG	435
Qy	580	TTTCTTTCTTTTGGGCTGGATGAGCAACTGTGTAGTGTCTGTGTATGGGAGCAACT	639
Db	434	TTTCTTTCTTTTGGGCTGGATGAGCAACTGTGTAGTGTCTGTGTATGGGAGCAACT	375
Qy	640	GGACGATGGGCGATTTTGTATCCAGAGATTTATCAACTTTTGTGTCAAACTAGGTTTGT	699
Db	374	GGACGATGGGCGATTTTGTATCCAGAGATTTATCAACTTTTGTGTCAAACTAGGTTTGT	315
Qy	700	GAGTGTACAGACCAAGCCATCATGCTCTGCTCTGGGATTCGAATGGGCCACCC	759
Db	314	GAGTGTACAGACCAAGCCATCATGCTCTGCTCTGGGATTCGAATGGGCCACCC	255
Qy	760	GGCTTTCACACTGCAGAAAGCCCTCCAGGGAGTTTACTGATAGTGTCTGAAGCTAAACTGAAGA	819
Db	254	GGCTTTCACACTGCAGAAAGCCCTCCAGGGAGTTTACTGATAGTGTCTGAAGCTAAACTGAAGA	195
Qy	820	GCCTGGATTTCCCTTTCTACTGATTTAAAGGATGGAAACTTGGAACTGGTAGCTAGT	879
Db	194	GCCTGGATTTCCCTTTCTACTGATTTAAAGGATGGAAACTTGGAACTGGTAGCTAGT	135
Qy	880	GCCTCTCTATCAATCAGGGTTGCACTTGGACATATCTACCTTAAATAGTGCATT	939
Db	134	GCCTCTCTATCAATCAGGGTTGCACTTGGACATATCTACCTTAAATAGTGCATT	75
Qy	940	TTTTCTCAGGAAATTTAGATGAATCAATTAATAACTCTCTAGCAAAATGAATCTCAAA	999
Db	74	TTTTCTCAGGAAATTTAGATGAATCAATTAATAACTCTCTAGCAAAATGAATCTCAAA	15
Qy	1000	TAA 1002	
Db	14	AAA 12	

RESULT 3

AK088859
LOCUS
DEFINITION

ACCESSION

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

AK088859 1403 bp mRNA linear HTC 20-SEP-2003
Mus musculus 2 days neonate thymic cells cDNA, RIKEN
full-length enriched library, clone:E430029E06
product:N-acetylneuraminase pyruvate lyase, full insert sequence.
AK088859 GI:26353975
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujikake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, T.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1403)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Benio, H., Carninci, P.,
Fukuda, S., Furum, M., Hanagaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imokani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanakawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saikou, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akita, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokotama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Adenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>

URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

source

Location/Qualifiers

1..1403

/organism="Mus musculus"

/mol_type="mRNA"

/strain="NOD"

/db_xref="FANTOM DB:E430c29E06"

/db_xref="MGI:2427756"

/db_xref="taxon:10090"

/clone="E430029E06"

/cell_type="thymic cells"

/tissue_type="thymus"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="2 days neonate"

86..1048

/note="unnamed protein product; N-acetylneuraminate

pyruvate lyase (MGI:1921341)

putative"

/codon_start=1

/protein_id="BAC40618.1"

/db_xref="GI:26353976"

/translation="MAPPKLRLGLVAATITPTWENGINFVIGYVDYLKSGQVK

NIFVNGTGGESLSVSRQVABEWNGRNKLDQVIVHVALNPKSDELQRAAE

IGAGVIAVAPFFFSQNDALISFLREVAAPLTPFYIHPMSMTGWKIRAEELLD

GQDKIPFQGLKFTDLDLDFGQVQDQHFQALLFGLVDSQLLSALVPMGTAVGS

TNYLKGKTNQMLEAFQDLASLSYQFRIQRFIVIKLGFVSGQTKAINTLVSGI

PMGPRLPLOKATQETAKAEAKLSLDFLSPSVKEGKPLASA"

1361..1386

/note="putative"

1403

/note="putative"

ORIGIN

Query Match 41.5%; Score 592.8; DH 11; Length 1403;
 Best Local Similarity 72.8%; Pred. No. 1.1e-157; Indels 114; Gaps 6;
 Matches 900; Conservative 0; Mismatches 222

QY 1 ATGGGCTTCCCAAGAAGAACTTCAGGGCTTGTGGCTGCAACCATCGCCAATGACT 60
 DB 86 ATGGGCTTCCCTTANGAAGAACTCCGGGGTCTGTGTGCTGCCACCATCACTCCAATGACA 145

QY 61 GAGAAATGAGAAATCACTTTTCACTAATGTGCTAGTATGTGATATCTTGTGAAGAA 120
 DB 146 GAGAAATGAGAAATCACTTTTCACTAATGTGCTAGTATGTGATATCTTGTGAAGAA 205

QY 121 CAGGAGTCAAGAAATTTTGTGATGCAACCAAGGAGAGGCTGTCCCTGAGCGTC 180
 DB 206 CAGGAGTCAAGAAATTTTGTGATGCAACCAAGGAGAGGCTGTCCCTGAGCGTC 265

QY 181 TCAGAGCGTCGCCAGGTTCAGAGAGTGGGTGACAAAAGGAGGAGCAAGCTGGATCAG 240
 DB 266 TCTGAACGCGCCAGCTCGCGAGGAATGGTTTACCAAGGAGGAGCAAGTTGGACCAAG 325

QY 241 GTGATTAATTCAGTAGGAGCTAGCTTGAAGAGTCAAGAACTGCCCCACATGCA 300
 DB 326 GTGGTGAATTCAGTGGAGCACTTAAACGTGAAGAGTCAAGAACTGCCCCACATGCA 385

QY 301 SCAGAAATAGGAGCTGATGCGCTGTGCTGCACTTGCACCGTCTTCTCAAGCCATGACC 360
 DB 386 GCAGAAATAGGAGCTGATGCGCTGTGCTGCACTTGCACCGTCTTCTTCAAGTCAAGAAC 445

QY 361 AAGATATCTCTGATTAATTTCTTAAGAGAGTGGGTGTGCTGCCGCCCTGCCCTGCCATTT 420

DB 446 AAGATGCCCCCTGATTAAGCTTCTCGAGGAGGTGGAGCTGCAGCCCTACTTTGCCATTT 505
 QY 421 TATTACTATCATATCTCTGCTTGCAGGGGTAAAGATTCTGTCTCAGGAGTTGTGGAT 480
 DB 506 TATTACTATCATATCTCTGCTTGCAGGGGTAAAGATTCTGTCTCAGGAGTTGTGGAT 565
 QY 481 GAGATCTCTGATTAAGTCCCACTTCCCAAGGGCTGAATTCAGTGATACAGATCTCTTA 540
 DB 566 GGGATTCAGGACCAAGATCCCACTTCCCAAGGGCTGAATTCAGTGATACAGATCTCTTA 625
 QY 541 GACTTCGGGCAATGTCTGATCAGAAATCGCCAGCAACAGATTTCTTCTTTTGGGGTG 600
 DB 626 GATTTCGGCCAATGTCTGATCAGAAATCACAGGAGACAGTTTCTTCTTTGGGGTG 665
 QY 601 GATGAGCAATCTGTGATGCTCTGTGATGAGGAGCAACTGGAGCAGTGGG----- 650
 DB 686 GACGACCAATCTGTGATGCTCTGTGATGAGGAGCAACTGGAGCAGTGGAGTACTAT 745
 QY 651 ----- 650
 DB 746 AACTACTGGGAAAAGAACCAACAGATGTGGAGGCTTTTGAACAGAGGACTTGGCT 805
 QY 651 -----CAGTTTGTATCCAGAGATTTTATCAACTTTGTGTTCAACAGTAGGT 695
 DB 806 TCAGCACTCAGTTACCAGTTTCGATCCAGAGATTTATCAACTATCTGATCAAGCTAGGT 865
 QY 696 TTTGAGTGTCAAGACCAAGCCATCATGACTCTGCTCTCTGGGATTTCCAAATGGGCCCA 755
 DB 866 TTTGAGTGTCAAGACCAAGCCATCATGACTCTCTCTCTGGGATTTCCAAATGGGCCCA 925
 QY 756 CCGGCTCTCCACTGCAGAAAGCCTCCAGGGAGTTTACTGATAGTGTGAAGCTAAACTG 815
 DB 926 CCGGCTCTCTCTGCGAAGAACCCAGGAGTTTACTGCTAAGCTGAGGCCAAGCTA 985
 QY 816 AAGAGCTGGATTTCTTTTCTCTGATTTAAAGGATGGAACCTTGAAGCTGTGTAGC 875
 DB 986 AAGAGCTGGATTTCTTTTCTCTGATTTAAAGGAGGGGAGGCTTTAGCTAGTGCC 1045
 QY 876 TAGTGTCTCTATCAATCAGGTTTGCACCTTGAAGACATATCTACCTTAATAGTGC 935
 DB 1046 TGAT-----AAGTCTGGGTGTGAAGCTTAGAC-TTATCTATCTTCAATAGTAC 1093
 QY 936 ATTTTCTCTCAGGAAATTTTAGATGAATCAATGAATCTCTCTAGCAATGAATCTC 995
 DB 1094 ATCTTCTCTCAGGAACTTTTGTAGTGAAT-TTGAACCCATCTTTTATTAGGAAGTA 1151
 QY 996 ACAATAAGCAATGAGTACCTTTGTGAGCTTAAAGCTTTATTTTGTGAGAGGGCAA 1055
 DB 1152 GAATCCCAACCAATCAATATTTGATTTATGAACCTTAAACTTTTGTGATGAAGAGTCA 1211
 QY 1056 AACTCTAGGAGTCAACTCTCAGTCATTTCATTTCCACAGATTTTGTGAGAGAAATTT 1115
 DB 1212 CA-----AGCCCTAAGTCATTCTTTGTA-AAATTTATAGAGAAATCT 1256
 QY 1116 CTGTTTATATGATGAATGAATGAATCAAGAGGAAAT 1151
 DB 1257 CTACTATACAGATGAATGAATGAATGAATGAAT 1292

RESULT 10

AK002734

LOCUS

DEFINITION

Mus musculus adult male kidney cDNA, RIKEN full-length enriched

library, clone:0610033B02 product:N-acetylneuraminate pyruvate

lyase, full insert sequence.

AK002734.1

GI:12832929

VERSION

HTC; CAP trapper.

MUS musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1


```

Db      602 TGTCTTTAG 610

RESULT 12
AV728536      753 bp mRNA linear EST 17-OCT-2000
DEFINITION   AV728536 HTC Homo sapiens cDNA clone HTCBBD11 5', mRNA sequence.
ACCESSION    AV728536
VERSION      AV728536.1 GI:10837957
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 753)
AUTHORS     Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
            Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L.,
            Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
            Chen, J., Chen, Z. and Han, Z.
TITLE       Homo sapiens cDNA HTC clones
JOURNAL     Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919 (ex. 45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.

FEATURES             Location/Qualifiers
     source           1..753
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HTCBBD11"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="HTC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"

ORIGIN
Query Match      37.6%; Score 537.2; DB 9; Length 753;
Best Local Similarity 95.0%; Pred. No. 6.7e-142;
Matches 587; Conservative 0; Mismatches 28; Indels 3; Gaps 3;

QY      667 GATTATCAACTTTGTTGTCCTCAACTAGTTTGGAGTGTCAAGACCAAGCCATCATGA 726
Db      90 GTTCTCTCACTTTGTTTCCGCTCTAGTTTGGAGTGTCAAGACCAAGCCATCATGA 149

QY      727 CTCTCGTCTCTGGGATTCCAATGGGCCACCCGGCTCCACTGCGAAGAGCTCCAGGG 786
Db      150 CTCTCGTCTCTGGGATTCCAATGGGCCACCCGGCTCCACTGCGAAGAGCTCCAGGG 209

QY      787 AGTTTACTGATAGTCTGAAGCTAACTGAAGAGCTGGATTCTCTTTCTCACTGATT 846
Db      210 AGTTTACTGATAGTCTGAAGCTAACTGAAGAGCTGGATTCTCTTTCTCACTGATT 269

QY      847 TAAGAGTGAAGAACTGAAGCTGGTAGTGTCTCTATCAAAATCAGGGTTTGCAC 906
Db      270 TAAGAGTGAAGAACTGAAGCTGGTAGTGTCTCTATCAAAATCAGGGTTTGCAC 329

QY      907 CTGAGACATAATCTACCTTAATAGTCATTTTTTCTCAGGGAATTTAGATGAATT 966
Db      330 CTGAGACATAATCTACCTTAATAGTCATTTTTTCTCAGGGAATTTAGATGAATT 389

QY      967 GAATAAACTCTCTAGCAAAATGAATCTCAAAATGAATGAGTACCTTTTGTGAGCC 1026
Db      390 GAACTAACTCTCTAGCAAAATGAATCTCAAAATGAATGAGTACCTTTTGTGAGCC 449

QY      1027 TTAATAAGCTTTATTTGTGAAGGGGAAAACCTCTAGGAGTCACAACTCTCACTTC 1086

```

```

Db      450 TTAATAAGCTTTATTTCTGAAGGGGCAAAAACCTTAGGAGTCACAACTCTCAGTCATTC 509
QY      1087 ATTTTCACAGATTTTTCGTGGAGAAATTTCTCTTATATGATGAAATCGAATCAAGAG 1146
Db      510 ATTTTCACAGATTTTTCGTGGAGAAATTTCTG-TTATATGGAATGAAATCGAATCAAGAG 568
QY      1147 AAAATTGTAATGATTAATTCATTCCTTTAGGAGCTCTCATTTATC-TCGGTCTCTGG 1205
Db      569 AAAATTGTAATGATTAATTCATTCCTCTAGGAGCTCTCATTTATCTTCTCTCTGG 628
QY      1206 TTCCTATTCCTATTATTAAGCTGCTTAATTTTAAACCACTATATATAT-GTCTTCATTTTA 1264
Db      629 GTTCTTAATCCTATTTTAAAGGTGGCTAATTTTAAACCACTATATATATGCTCTCAATTTTA 688
QY      1265 ATAATAATTCATTTGGAA 1282
Db      689 ATAATAATTCATTTGGA 706

RESULT 13
BU658094      643 bp mRNA linear EST 30-SEP-2002
LOCUS        BU658094
DEFINITION   BU658094 c132a06.z1 Hembase: Erythroid Precursor Cells (LCB:cl library) Homo
ACCESSION    BU658094
VERSION      BU658094.1 GI:23370276
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 643)
AUTHORS     Gubin,A.N., Lee,Y.T., Bouffard,G.G. and Miller,J.L.
TITLE       Gene Expression in Human Erythroid Precursor Cells
JOURNAL     Unpublished (2002)
COMMENT     Contact: Jeffery L. Miller
            Laboratory of Chemical Biology
            National Institute of Diabetes and Digestive and Kidney Diseases
            Building 10, Room 9B17, National Institutes of Health, Bethesda, MD
            20892, USA
            Tel: 301 402 2373
            Fax: 301 435 5148
            Email: jlm7@nih.gov
            The 'cl' library was constructed by Alexander Gubin, Ph.D. in the
            Laboratory of Chemical Biology, NIDDK, NIH. DNA Sequencing and/or
            analyses by National Institutes of Health Intramural Sequencing
            Center (NISC). More information available at:
            http://hembase.niddk.nih.gov
            Plate: 32 row: a column: 06
            Seq primer: 5' lambda-Triplex2 Sequencing Primer.

FEATURES             Location/Qualifiers
     source           1..643
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="c132a06"
                     /sex="unknown"
                     /tissue_type="blood"
                     /cell_type="Erythroid Precursor Cells"
                     /cell_line="Primary Culture of Peripheral Blood
                     Mononuclear Cells"
                     /dev_stage="Precursor erythroblasts; GPA++"
                     /lab_host="DH5alpha"
                     /clone_lib="Hembase; Erythroid Precursor Cells (LCB:cl
                     library)"
                     /note="Organ: blood; Vector: pTriplex2; Site 1: SfiI;
                     Site 2: SfiI; A complementary DNA (cDNA) library from
                     human erythroid precursor cells was constructed using
                     SMART PCR (polymerase chain reaction) cDNA library
                     Construction Kit (Clontech, Palo Alto, CA) according to
                     the manufacturer's directions, but with slight
                     modifications. Briefly, reverse transcription was
                     performed in the presence of 1 umol/L peptide nucleic acid

```

(PNA) oligos
(N-terminal)-biotin-GTC-CAC-CCG-AAG-CTT-G-(C-terminal) and
(N-terminal)-biotin-C(T/C)T-GAA-GTT-CTC-AGG-A-(C-terminal)
- Synthesized cDNA was digested with SfiI and
size-selected on a 1% agarose gel (>800bp). Large-scale
sequencing of the library was performed by the NIH
Intramural Sequencing Center (NISC;
Http://www.nisc.nih.gov/).

ORIGIN
Query Match 37.2%; Score 531.4; DB 13; Length 643;
Best Local Similarity 99.1%; Pred. No. 2.8e-140;
Matches 532; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCAAGAGAACTTCAGGCTTGTGGTGCACCCNTCACCCCAATGACT 60
DB 111 ATGGCCCTCCCAAGAGAACTTCAGGCTTGTGGTGCACCCCAATGACT 170
QY 61 GAGAAATGAGAAATCACTTTCAGTAAATGGTCAATGATGGATTATCTTGTGAAGAA 120
DB 171 GAGAAATGAGAAATCACTTTCAGTAAATGGTCAATGATGGATTATCTTGTGAAGAA 230
QY 121 CAGGAGTGAAGACATTTTGTGAATGSCACACAGAGAGAGCTGTCCCTGAGGCTC 180
DB 231 CAGGAGTGAAGACATTTTGTGAATGSCACACAGAGAGAGCTGTCCCTGAGGCTC 290
QY 181 TCAGAGGCTGCCAGAGTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 240
DB 291 TCAGAGGCTGCCAGAGTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 350
QY 241 GTGTAATTCAGTAGAGACTGAGCTTGAAGAGTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 300
DB 351 GTGTAATTCAGTAGAGACTGAGCTTGAAGAGTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 410
QY 301 GCAGAAATAGGAGCTGATGAGTGCATGCTGATTCAGAGGCTGTCCCTCAAGCCATGAGAC 360
DB 411 GCAGAAATAGGAGCTGATGAGTGCATGCTGATTCAGAGGCTGTCCCTCAAGCCATGAGAC 470
QY 361 AAAGATATCCTGATTAATTTCTTAAAGGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 420
DB 471 AAAGATATCCTGATTAATTTCTTAAAGGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 530
QY 421 TATTACTATCATCTCCCTGACAGGGTGAAGATTCGTGCTGAGAGTGTGATGAT 480
DB 531 TATTACTATCATCTCCCTGACAGGGTGAAGATTCGTGCTGAGAGTGTGATGAT 590
QY 481 GGGATTCGTGATAGATCCCAACCTTCAGAGGCTGAAATTCAGTGATACAGA 533
DB 591 GGGATTCGTGATAGATCCCAACCTTCAGAGGCTGAAATTCAGTGATACAGA 643

RESULT 14
CB999084
LOCUS
DEFINITION
AGENCOURT 13631478 NIH_MGC.186 Homo sapiens cDNA clone
IMAGE:30325020 5', mRNA sequence.
ACCESSION
CB999084
VERSION
CB999084.1 GI:30325020
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGNISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 893)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM130 row: c column: 13
High quality sequence stop: 560.
Location/Qualifiers
1-893
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30325020"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC.186"
/note="Organ: Pooled-Skin; Vector: pDNR-LIB; Site 1: SfiI
(ggccattatggcc); Site 2: SfiI (ggcgcctggcc); Library is
oligo-dt primed and directionally cloned. cDNA was
prepared from a pooled samples of tissues from Skin,
meninges, duramater, pia matter and choroid plexus.
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCGCATG-dt(30)BN-3'
sequence: 5'-ATTCTAGAGCGCGGCGGCGGCGCATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library"

ORIGIN

Query Match 37.2%; Score 531; DB 14; Length 893;
Best Local Similarity 99.1%; Pred. No. 4.3e-140;
Matches 534; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGGCCCTCCCAAGAGAACTTCAGGCTTGTGGTGCACCCCAATGACT 60
DB 106 ATGGCCCTCCCAAGAGAACTTCAGGCTTGTGGTGCACCCCAATGACT 165
QY 61 GAGAAATGAGAAATCACTTTCAGTAAATGGTCAATGATGGATTATCTTGTGAAGAA 120
DB 166 GAGAAATGAGAAATCACTTTCAGTAAATGGTCAATGATGGATTATCTTGTGAAGAA 225
QY 121 CAGGAGTGAAGAACATTTTGTGAATGSCACACAGAGAGAGCTGTCCCTGAGGCTC 180
DB 226 CAGGAGTGAAGAACATTTTGTGAATGSCACACAGAGAGAGCTGTCCCTGAGGCTC 285
QY 181 TCAGAGCTGCCAGAGTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 240
DB 286 TCAGAGCTGCCAGAGTGCAGAGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 345
QY 241 GTGTAATTCAGTAGAGCACTGAGCTTGAAGAGTGCACAGAGAACTGGCCCAATGCA 300
DB 346 GTGTAATTCAGTAGAGCACTGAGCTTGAAGAGTGCACAGAGAACTGGCCCAATGCA 405
QY 301 GCAGAAATAGGAGCTGATGAGTGCATGCTGATTCAGCGCTTCTTCTGAGCGATGAC 360
DB 406 GCAGAAATAGGAGCTGATGAGTGCATGCTGATTCAGCGCTTCTTCTGAGCGATGAC 465
QY 361 AAAGATATCCTGATTAATTTCTTAAAGGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 420
DB 466 AAAGATATCCTGATTAATTTCTTAAAGGAGTGGTGTGACAAAGGAGGACCAAGCTGGATCAG 525
QY 421 TATTACTATCATCTCCCTGACAGGGTGAAGATTCGTGCTGAGGATTTGTTGGAT 480
DB 526 TATTACTATCATCTCCCTGACAGGGTGAAGATTCGTGCTGAGGATTTGTTGGAT 585
QY 481 GGGATTCGTGATAGATCCCAACCTTCAGAGGCTGAAATTCAGTGATACAGATCCTCT 539
DB 586 GGGATTCGTGATAGATCCCAACCTTCAGAGGCTGAAATTCAGTGATACAGATCCTCT 644

RESULT 15
CB985894
LOCUS

CB985894 723 bp mRNA linear EST 01-MAY-2003

DEFINITION IMAGE:3025735 NIH_MGC_184 Homo sapiens cDNA clone
 IMAGE:3025735 5', mRNA sequence.
 CB95894
 VERSION CB95894.1 GI:30280418
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 723)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILLNL at:
<http://image.llnl.gov>
 Plate: NDCM132 row: a column: 38
 High quality sequence stop: 558.
 Location/Qualifiers
 1..723
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3025735"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_184"
 /note="Organ: Pooled-Glandular; Vector: pDNR-LIB; Site 1:
 SfiI (ggccattatggcc); Site 2: SfiI (ggccgctcggcc);
 Library is oligo-dT primed and directionally cloned. cDNA
 was prepared from a glandular pool of tissues from thyroid,
 parathyroid, adrenal, cortex and pineal gland. 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGAGGCGCGCATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.38
 kb (range 0.60-3.5 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC library."

ORIGIN

Query Match 36.9%; Score 528; DB 14; Length 723;
 Best Local Similarity 98.8%; Pred. No. 2.8e-139;
 Matches 553; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY	1	ATGGCCTTCCCAAGAGAAGAACTTCAGGGTCTTGGCTGCAACCATCAGCCCAATGACT	60
DB	108	ATGGCCTTCCCAAGAGAAGAACTTCAGGGTCTTGGCTGCAACCATCAGCCCAATGACT	167
QY	61	GAGAAATGGAGAAATCACTTTTCAGTAATGGTCAGTATGTTGATTTCTTGTAAGAA	120
DB	168	GAGAAATGGAGAAATCACTTTTCAGTAATGGTCAGTATGTTGATTTCTTGTAAGAA	227
QY	121	CAGGGAGTGAAGAACATTTTGTGATGGCAACACAGAGAGAGCCCTGTCCCTGAGCGTC	180
DB	228	CAGGGAGTGAAGAACATTTTGTGATGGCAACACAGAGAGAGCCCTGTCCCTGAGCGTC	287
QY	181	TCAGAGCGTCGCCAGGTTGCAGAGGAGTGGGTGACAAAAGGAAGGACAAGCTGGATCAG	240
DB	288	TCAGAGCGTCGCCAGGTTGCAGAGGAGTGGGTGACAAAAGGAAGGACAAGCTGGATCAG	347
QY	241	GTGATTAATTCAGCTAGGAGCACTGAGCTTGAAGAGTTCACAGGAATGGCCCAACATGCA	300
DB	348	GTGATTAATTCAGCTAGGAGCACTGAGCTTGAAGAGTTCACAGGAATGGCCCAACATGCA	407
QY	301	GCAGAAATAGGAGCTGATGGCATCGCTGTCAATGACCCGTTCTTCTCAAGCCATGAC	360

Db	408	GCAGAAATAGGAGCTGATGGCATCGCTGTCAATGACCCGTTCTTCTCAAGCCATGAC	467
QY	361	AAAGATATCCTGATTAATTTCTTAAAGGAAGTGGTGTCTGGCCGCCCTGCCCTGCCATTT	420
Db	468	AAAGATATCCTGATTAATTTCTTAAAGGAAGTGGTGTCTGGCCGCCCTGCCCTGCCATTT	527
QY	421	TATTAATATCACTTCCTGCTGTGACAGGGGTAAAGATTCGTGCTGAGGAGTGTGGAT	480
Db	528	TATTAATATCACTTCCTGCTGTGACAGGGGTAAAGATTCGTGCTGAGGAGTGTGGAT	587
QY	481	GGGATTCCTGGATAAGATCCGACCTT-CCAGGGGCTGAAATTCAGTGATACAGAT-CTCT	538
Db	588	GGGATTCCTGGATAAGATCCGACCTT-CCAGGGGCTGAAATTCAGTGATACAGAT-CTCT	647
QY	539	TAGACTTCGGGCAATGTGTT	558
Db	648	TAACTTCGGGCAATGTGTT	667

Search completed: May 11, 2004, 11:45:32
 Job time : 4160.87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:29 ; Search time 109.968 Seconds
(without alignments)
7211.402 Million cell updates/sec

Title: US-09-930-440B-1

Perfect score: 1429

Sequence: 1 atggccttcctccaaagaagaa.....cttttgattcatttcgatg 1429

Scoring table: IDENTITY NUC
Gapop 1C.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1429	100.0	1429	4	US-09-516-143A-5
2	1332.4	93.2	2536	2	US-09-027-013-2
3	1332.4	93.2	2536	3	US-09-244-233-2
4	792	55.4	1923	4	US-09-833-381-1100
5	654	45.8	1127	4	US-09-833-381-1099
6	68.4	4.8	903	4	US-09-543-681A-2696
7	51.4	3.6	927	1	US-09-543-681A-254
8	51.4	3.6	7218	1	US-08-232-463-14
9	44.2	3.1	1368	3	US-08-874-563-5
10	44.2	3.1	1368	3	US-08-577-483-14
11	44.2	3.1	4253	3	US-08-577-483-7
12	43.2	3.0	9897	4	US-08-561-527-10
13	41.8	2.9	4254	2	US-08-443-639-7
14	41.8	2.9	4254	4	US-09-435-380-7
15	41.8	2.9	23439	4	US-08-556-171B-38
16	39.8	2.8	1664976	4	US-08-916-421B-1
17	39.6	2.7	4239	4	US-08-956-171B-333
18	39.2	2.8	832	4	US-09-621-976-2813
19	38.8	2.7	5181	1	US-08-257-073-10
20	38.6	2.7	832	4	US-09-621-976-2813
21	38.6	2.7	640881	4	US-09-790-988-1
22	38.4	2.7	99916	4	US-09-816-095-3
23	38	2.7	1128	4	US-09-710-000-9
24	37.8	2.6	128779	4	US-09-497-855A-38
25	37.6	2.6	1135	4	US-08-936-165A-135
26	37.6	2.6	640681	4	US-09-790-988-1
27	37.2	2.6	128779	4	US-09-497-855A-38

C	28	37	2.6	1219	4	US-08-195-705-1	Sequence 1, Appli
	29	36.8	2.6	6831	2	US-08-609-049A-27	Sequence 27, Appl
	30	36.8	2.6	6831	3	US-09-170-996-27	Sequence 27, Appl
	31	36.8	2.6	8961	4	US-13-204-708-80	Sequence 80, Appl
C	32	36.4	2.5	2614	4	US-09-004-056-1	Sequence 1, Appli
	33	36.4	2.5	5340	4	US-09-627-122-21	Sequence 21, Appl
	34	36.4	2.5	9636	1	US-08-323-170B-1	Sequence 1, Appli
C	35	36.4	2.5	9636	4	US-08-954-441-1	Sequence 1, Appli
	36	36	2.5	5152	4	US-09-833-381-831	Sequence 831, App
	37	36	2.5	5152	4	US-10-204-708-73	Sequence 73, App
	38	35.8	2.5	24979	2	US-08-147-777-3	Sequence 3, Appli
	39	35.8	2.5	24979	3	US-08-452-872-3	Sequence 3, Appli
	40	35.8	2.5	24979	5	PCT-US93-03885-3	Sequence 59, Appl
C	41	35.4	2.5	6156	4	US-10-204-708-59	Sequence 3, Appli
	42	35.4	2.5	65042	4	US-09-784-316-3	Sequence 107, App
	43	35.2	2.5	6405	4	US-08-961-527-107	Sequence 72, Appl
	44	35.2	2.5	8607	4	US-10-204-708-72	Sequence 11, Appl
C	45	35	2.4	9620	3	US-08-952-127-11	

ALIGNMENTS

RESULT 1
US-09-516-143A-5
; Sequence 5, Application US/09516143A
; Patent No. 633482

; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF050PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; CURRENT FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1429
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(690)
US-09-516-143A-5

Query Match	100.0%;	Score 1429;	DB 4;	Length 1429;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 1429;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	ATGGCCTTCCCAAGAGAAACTTCAGGCTCTGTGGCTGCAACCATCATCGCCAAATGACT	60		
Db	1	ATGGCCTTCCCAAGAGAAACTTCAGGCTCTGTGGCTGCAACCATCATCGCCAAATGACT	60		
QY	61	GAGAATGGAATAATCAACTTTTCAGTAAATGGTTCAGTATGTGATTATCTTGTGAAGAA	120		
Db	61	GAGAATGGAATAATCAACTTTTCAGTAAATGGTTCAGTATGTGATTATCTTGTGAAGAA	120		
QY	121	CAGGAGTAGGAACATTTTGTGAATGGCAACACAGGAGAGCCCTGTCCCTGAGCGTC	180		
Db	121	CAGGAGTAGGAACATTTTGTGAATGGCAACACAGGAGAGCCCTGTCCCTGAGCGTC	180		
QY	181	TCAGAGCGTCGCCAGGTTGCAGAGAGTGGGTGCACAAAAGGAGACAAAGCTGGATCAG	240		
Db	181	TCAGAGCGTCGCCAGGTTGCAGAGAGTGGGTGCACAAAAGGAGACAAAGCTGGATCAG	240		
QY	241	GTGATAATTCACGTAGGACACTGAGCTTCAAGGAGTCAACGAAATGGCCCCAACATGCA	300		
Db	241	GTGATAATTCACGTAGGACACTGAGCTTCAAGGAGTCAACGAAATGGCCCCAACATGCA	300		
QY	301	CGAGAAATAGGAGCTGATGGCATCGCTGTCAATGCAACCGTCTTCTCTCAAGCCATGGACC	360		
Db	301	CGAGAAATAGGAGCTGATGGCATCGCTGTCAATGCAACCGTCTTCTCTCAAGCCATGGACC	360		

361 AAAGATATCTGATTAATTTCTTAAGGAAGTGGCTGTGCGCCCTGCGCTGCATTT 420
 Db AAAGATATCTGATTAATTTCTTAAGGAAGTGGCTGTGCGCCCTGCGCTGCATTT 420
 Qy TAATTAATCAATCTGCTGCTGATGAGGGGTAAGATTCGTGCTGAGGATTTGGAT 480
 Db TAATTAATCAATCTGCTGCTGATGAGGGGTAAGATTCGTGCTGAGGATTTGGAT 480
 Qy GGGATTCGGATAGATCCCACTTCCCAAGGGCTGAAATTCAGTGATACAGATCTCTTA 540
 Db GGGATTCGGATAGATCCCACTTCCCAAGGGCTGAAATTCAGTGATACAGATCTCTTA 540
 Qy GACTTCGGGCAATGTTGATCAGATCCGAGACAGATTTGCTTCTCTTTGGGGTG 600
 Db GACTTCGGGCAATGTTGATCAGATCCGAGACAGATTTGCTTCTCTTTGGGGTG 600
 Qy GATGAGCAATCTGCTGAGTCTGCTGATGGAGCAACTGGAGCACTGGGCACTTTTGT 660
 Db GATGAGCAATCTGCTGAGTCTGCTGATGGAGCAACTGGAGCACTGGGCACTTTTGT 660
 Qy TCCAGAGATTTATCAACTTTGTTGTCAACTAGGTTTGGAGTGCACAGACCAAGCCA 720
 Db TCCAGAGATTTATCAACTTTGTTGTCAACTAGGTTTGGAGTGCACAGACCAAGCCA 720
 Qy TCATGACTCTGCTCTGAGATTCGAATGGGCGCCACCCCGCTTCCACTGCAGAAAGCT 780
 Db TCATGACTCTGCTCTGAGATTCGAATGGGCGCCACCCCGCTTCCACTGCAGAAAGCT 780
 Qy CCAGGAGTTTACTGATGCTGAGCTTAAGCTTAAGCTGAAGCACTGGATTTCTTTTCA 840
 Db CCAGGAGTTTACTGATGCTGAGCTTAAGCTGAAGCACTGGATTTCTTTTCA 840
 Qy CTGATTTAAGGATGGAATCTTGGAGCTGCTAGCTAGTGCCTCTCTATCAAAATCAGGCT 900
 Db CTGATTTAAGGATGGAATCTTGGAGCTGCTAGCTAGTGCCTCTCTATCAAAATCAGGCT 900
 Qy TTGCACCTTGACATATCTACCTTAATAGTGAATTTTCTCAGGGATTTTATAGT 960
 Db TTGCACCTTGACATATCTACCTTAATAGTGAATTTTCTCAGGGATTTTATAGT 960
 Qy GAACTTGAATAAATCTCTAGCAATGAAATCTCACAATAAGCACTGAGTACCTTTT 1020
 Db GAACTTGAATAAATCTCTAGCAATGAAATCTCACAATAAGCACTGAGTACCTTTT 1020
 Qy TGAGCTTTAAGATCTTAATTTTGGAGGAGAAATTTCTGTTATATGATGAATGGAATC 1080
 Db TGAGCTTTAAGATCTTAATTTTGGAGGAGAAATTTCTGTTATATGATGAATGGAATC 1080
 Qy TCATTCATTTACAGATTTTCTGAGGAAATTTCTGTTATATGATGAATGGAATC 1140
 Db TCATTCATTTACAGATTTTCTGAGGAAATTTCTGTTATATGATGAATGGAATC 1140
 Qy AAGAGGAAATTTGTAATTAATTCATCTGCTTTAGGAGCTCTCATATCTCGGTC 1200
 Db AAGAGGAAATTTGTAATTAATTCATCTGCTTTAGGAGCTCTCATATCTCGGTC 1200
 Qy TCTGTTTCTATCTATTTTAAAGTGTCTTAATTTTAAACCACTATAATGTTCTCAT 1260
 Db TCTGTTTCTATCTATTTTAAAGTGTCTTAATTTTAAACCACTATAATGTTCTCAT 1260
 Qy TTTAATAATATTTCAATTTAGGAAATCTGAGCTACTGCTGCTAGGAGGAC 1320
 Db TTTAATAATATTTCAATTTAGGAAATCTGAGCTACTGCTGCTAGGAGGAC 1320
 Qy TTTAATAACCAATCTGAACTCTCACTGATACAACTCAAAATPACACCACTCATTT 1380
 Db TTTAATAACCAATCTGAACTCTCACTGATACAACTCAAAATPACACCACTCATTT 1380
 Qy GSCTGCTAGTCTAATCTAGATGATGATCTTTTGAATTCATTTGATG 1429
 Db GSCTGCTAGTCTAATCTAGATGATGATCTTTTGAATTCATTTGATG 1429

RESULT 2
 US-09-027-013-2
 ; Sequence 2, Application US/09027013
 ; Patent No. 5962302
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Corley, Neil C.
 ; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/027,013
 ; FILING DATE: Herewith
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0462 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-855-0555
 ; TELEFAX: 650-845-4166
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2536 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: BBSNOT07
 ; CLONE: 2125668
 ; US-09-027-013-2

Query Match 93.2%; Score 1332.4; DB 2; Length 2536;
 Best Local Similarity 94.3%; Pred. No. 0;
 Matches 1428; Conservative 0; Mismatches 1; Indels 85; Gaps 1;
 Qy 1 ATGGGCTTCCCAAGAGAACTTCAAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 60
 Db 89 ATGGGCTTCCCAAGAGAACTTCAAGGCTCTGTGGCTGCAACCATCAGCCCAATGACT 145
 Qy 61 GAGATGAGAAATCAACTTTTCAATTAATGCTGATGATGATGATGATGATGATGATGAT 120
 Db 149 GAGATGAGAAATCAACTTTTCAATTAATGCTGATGATGATGATGATGATGATGATGAT 208
 Qy 121 CAGGAGTGAAGAACTTTTCTGAATGCAACAAGGAGAGGCTGCTGCTGAGCGTC 180
 Db 209 CAGGAGTGAAGAACTTTTCTGAATGCAACAAGGAGAGGCTGCTGCTGAGCGTC 268
 Qy 181 TCAGAGCTGCGCCAGGTTGCAAGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
 Db 269 TCAGAGCTGCGCCAGGTTGCAAGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 328
 Qy 241 GTGATAATTCAGTATGAGGAGCTGAGCTTGAAGGAGTCAAGGAACTGCGCCCAATGCA 300
 Db 329 GTGATAATTCAGTATGAGGAGCTGAGCTTGAAGGAGTCAAGGAACTGCGCCCAATGCA 388
 Qy 301 GCAGAAATAGGAGCTGATGGCATCGCTGCTGATGCAACCGTTCTTCTCAAGCCATGGAC 360

[illegible]

QY 865 AAGCTGCTAGCTAGTGGCTCTCTATCAAAATCAGGGTTTGCACCTTGGACATAATCTACC 924
Db 1503 AAGCTGCTAGCTAGTGGCTCTCTATCAAAATCAGGGTTTGCACCTTGGACATAATCTACC 1444
QY 925 TTAATAGTGCATTTTCTCAGGAAATTTAGATGAACCTTGAAATAACCTCTCTAGCA 984
Db 1443 TTAATAGTGCATTTTCTCAGGAAATTTAGATGAACCTTGAAATAACCTCTCTAGCA 1384
QY 985 AATGAATCTCACAAATAGCAATTCAGGTACCTTTTGTGAGCCTTAAAGTCTTAATTTG 1044
Db 1383 AATGAATCTCACAAATAGCAATTCAGGTACCTTTTGTGAGCCTTAAAGTCTTAATTTG 1324
QY 1045 TGAAGGGCAAAAATCTTAGGAGTCAACAATCTCAGTCATTCATTCACAGATTTTGTG 1104
Db 1323 TGAAGGGCAAAAATCTTAGGAGTCAACAATCTCAGTCATTCATTCACAGATTTTGTG 1264
QY 1105 TGGAGAAATTTCTCTTATATGGATGAATGGAATCAAGAGAAATTTGTAATGATTA 1164
Db 1263 TGGAGAAATTTCTCTTATATGGATGAATGGAATCAAGAGAAATTTGTAATGATTA 1204
QY 1165 TCCCATCTGCTTAGGAGTCTCATATATCTCGGTCTCTGTTCTCTATCTTAATTTTAA 1224
Db 1203 TCCCATCTGCTTAGGAGTCTCATATATCTCGGTCTCTGTTCTCTATCTTAATTTTAA 1144
QY 1225 GTTCTCTAATTTTAAACCTATATATATGCTCTCAATTTTAAATAATTTCAATTTGGAATC 1284
Db 1143 GTTCTCTAATTTTAAACCTATATATATGCTCTCAATTTTAAATAATTTCAATTTGGAATC 1084
QY 1285 TAGAAGAACTCTGAGTCTACTGATTTAGGAGGCACTTTAATACAACTCTACATGTC 1344
Db 1083 TAGAAGAACTCTGAGTCTACTGATTTAGGAGGCACTTTAATACAACTCTACATGTC 1024
QY 1345 TCAACTGTATACAACTCAAAATACACAGCTCATTTGCTGCTCAGTCACTAATCTAGAT 1404
Db 1023 TCAACTGTATACAACTCAAAATACACAGCTCATTTGCTGCTCAGTCACTAATCTAGAT 964
QY 1405 -GGATGCTTTGAAATTCATTTGGATG 1429
Db 963 GGGATGCTTTGAAATTCATTTGGATG 938

RESULT 5

US-09-833-381-1099
; Sequence 1099, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1099
; LENGTH: 1127
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1099

Query Match 45.8%; Score 654; DB 4; Length 1127;
Best Local Similarity 100.0%; Pred. No. 8.2e-186;
Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCAACCATCAGGCAATGACT 60
Db 199 ATGGCCCTCCCAAGAGAACTTCAGGCTCTTGGCTGCAACCATCAGGCAATGACT 258
QY 61 GAGAAATGAGAAATCAACTTTTCACTAATTTGGTCACTATGCTGATTTCTTGAAGAA 120
Db 259 GAGAAATGAGAAATCAACTTTTCACTAATTTGGTCACTATGCTGATTTCTTGAAGAA 318

QY 121 CAGGAGCTGAAGAACATTTTCTGAATGGCAACAACAGAGAGAGGCTCTCTCTGAGCGTC 180
Db 319 CAGGAGCTGAAGAACATTTTCTGAATGGCAACAACAGAGAGAGGCTCTCTCTGAGCGTC 378
QY 181 TCAGAGCCTCGCCACAGTTTGCAGAGGAGTGGGTGACAAAGAGGAGCAAGCTGGATCAG 240
Db 379 TCAGAGCCTCGCCACAGTTTGCAGAGGAGTGGGTGACAAAGAGGAGCAAGCTGGATCAG 438
QY 241 GTGATAATTCAGTATAGGAGCACTGAGCTTGAAGAGTCAACAGAACTGSCCCCAATGCA 300
Db 439 GTGATAATTCAGTATAGGAGCACTGAGCTTGAAGAGTCAACAGAACTGSCCCCAATGCA 498
QY 301 GCAGAAATAGAGCTGATGGCACTCGCTGTCATTGACCGTCTTCTCTAGAGCATGGACC 360
Db 499 GCAGAAATAGAGCTGATGGCACTCGCTGTCATTGACCGTCTTCTCTAGAGCATGGACC 558
QY 361 AAAGATATCTGATTAATTTCTTAAAGGAGTGGCTGCTGCGGCCCTGCCCCCATTT 420
Db 559 AAAGATATCTGATTAATTTCTTAAAGGAGTGGCTGCTGCGGCCCTGCCCCCATTT 618
QY 421 TATTACTATACATTTCTGCTTGAAGAGTAAAGATTCGCTGAGAGTTTGGAT 480
Db 619 TATTACTATACATTTCTGCTTGAAGAGTAAAGATTCGCTGAGAGTTTGGAT 678
QY 481 GGGATTTCTGATAAGATCCCACTTCCAGGCTGAAATTCAGTATACAGATCTCTTA 540
Db 679 GGGATTTCTGATAAGATCCCACTTCCAGGCTGAAATTCAGTATACAGATCTCTTA 738
QY 541 GACTTGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTTGTCTTCTTTTGGGCTG 600
Db 739 GACTTGGGCAATGTGTGATCAGAAATCGCCAGCAACAGTTTGTCTTCTTTTGGGCTG 798
QY 501 GATCAGCAACTGTGTGAGTGTCTGCTGATGGAGCAACTGAGCAGTGGGCACT 654
Db 799 GATCAGCAACTGTGTGAGTGTCTGCTGATGGAGCAACTGAGCAGTGGGCACT 852

RESULT 5

US-09-543-681A-2696
; Sequence 2696, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2696
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2696

Query Match 4.8%; Score 68.4; DB 4; Length 903;
Best Local Similarity 49.3%; Pred. No. 2.1e-10;
Matches 267; Conservative 0; Mismatches 266; Indels 9; Gaps 3;

QY 13 AAGAAGAAATTCAGAGTCTTGTGCTGCAACATCAGCGCAATGATGAGATGAGAA 72
Db 10 ATGAACAAATCTCTGAGCTGATTCGCGGCTCTACTACTCTTTTCTGCGGAGTAGT 69
QY 73 ATCAACTTTTCAGTAAATTCAGTATGTSATTAATCTTGTGAAGAACAGGAGTGAAG 132
Db 70 GTTAATTAACCTGTTATTG---ATGATATGCTAAACACTAATTAAGACTGGGTAACA 126
QY 133 AACATTTTGTGAATGGCAACAAGAGAGGCTCTCTCTGAGCGCTCTCAGAGCGTGGC 192
Db 127 GGGGCTTATGCTCTAGTACTACTGCTGAAGGTATACATTTGCTCGGTTGAAGAGGTAAA 186

193	QY	CAGGTTCCAGAGGAGTGGGTGACAAAAAGGAAGACAAAGCTGGATCAGGTGATGATTCAC	252
187	Db	AAAGTTGCTGAACGTTGGTTACAGCGAGTCAGGAGCAACTTGATC---TTATTATTTCAT	243
253	QY	GTAGGAGCACTGAGCTTTGAAGGAGTCAAGAGACTGSCCCCAACATGACGACGAGAAATAGA	312
244	Db	ACTGGTGCATTAAATATTGCTGATACGTAGAACTTGCTCGTCATGCTGAACACTTAGAT	303
313	QY	GCTGATGGCATCGCTGTCATTGSCACCGTCTTTCTTCAAGCCATGACCAAGATATCCCTS	372
304	Db	ATTAAAGCCACATCAGTGAATGGCCCTTCCTTTTAAACCTAGTCACTGTCATGATGA	363
373	QY	ATTAATTTCTTAAAGAAAGTGGCTGCTCGCGGCCCTCGCCCTCCCATTTTATTACTATCAC	432
364	Db	GTACATATTTCTGTTTAGCAGCGCAAGTCACCCCTCAAAAGGTTTCTACTATTATTCAC	422
433	QY	ATTCCCTGCCCTTGACAGGGGTAAAGATTGTCGTGAGGAGTTCGTTGGATGGGACTCTGGAT	492
423	Db	--TTCCACCACGTGACGATTAAAGCATTGATATGGAAAGTTTTTACAGACACAGGTAA	480
493	QY	AAGATCCCCACCTTTCCAGGGGCTGAAATTCAGTGATACAGATCTCTTAGACTTCGGGCAA	552
481	Db	GTGATCCCCCAATTTATCTGGCATGAAATTAACCTCCCTGATATGATGTATGTGAAATTCAGCGT	540
553	QY	TG 554	
541	Db	TG 542	

RESULT 7
 US-09-543-681A-254
 ; Sequence 254, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,756
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 254
 ; LENGTH: 527
 ; TYPE: DNA
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-254

QY 405 CCCGCGCTGCCATTTTATTACTATCACATTCC 437
||| | ||| ||| ||| |||
DP 402 AGAGAAATTCGCCAATCTATCTTTTATAACATTCC 434

RESULT \$
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4139
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: pTZgpt-fls
US-08-232-463-14

Query Match: 3.6%; Score 51.4; DB 4; Length 927;
Best Local Similarity 49.2%; Pred. No. 2.7e-05;
Matches 164; Conservative 0; Mismatches 166; Idels 3; Gaps 1;

	QY	252 CGTAGGAGCACTGAGCTTGAAGGAGTCACAGGAACTGGCCCAACATGCACGAGAATAAGG	311
	Dδ	1142 RR	1083
	QY	312 AGCTGATGGCATCGCTGTCAATTGCACCGTTCTTCCTCAAGCCATGGACCACAAGATATCCT	371
	Dδ	1082 RR	1023
	QY	372 G 372	
	Dδ	1022 G 1022	

RESULT 9

```

US-08-874-563-5/c
; Sequence 5, Application US/08874563
; Patent No. 6002071
; GENERAL INFORMATION:
; APPLICANT: CHAPPELL, JOSEPH
; APPLICANT: NEWMAN, JEFFREY D.
; APPLICANT: YIN, SHAOHUI
; TITLE OF INVENTION: TRANSCRIPTIONAL SILENCING
; TITLE OF INVENTION: ELEMENTS AND THEIR BINDING FACTORS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,563
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/020,087
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark, Esq.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 07678/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1368 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-874-563-5

```

	Query Match	3.1%;	Score 44.2;	DB 3;	Length 1358;
	Best Local Similarity	47.1%;	Pred. No. 0.0048;		
	Matches 136;	Conservative 0;	Mismatches 153;	Indels 0;	Gaps 0;
QY	1007	TGAGGTACCTTTTGTGAGCCTTAAAGTCTATTTTTGTGAGGGGCAAAAACCTCAGGA	1066		
DB	437	TGACATAGTGATAATGTATCTTTACGCCCTTTATTTCACGAAAAAAAAGTTGCATGG	378		
QY	1067	GTCACAACCTCTCAGTCATTCATTTCCAGAGTTTTTTGTGGAGAAAATTCCTGTTTATATG	1126		
DB	377	TCGGCGGATCGGATACACACACACACATCATTTATATATATATATATATATA	318		
QY	1127	GATGAAATGGAAATCAAGAGGAAAAATGTAAATCGATAATTCATCTCTCTTTAGGAGCTC	1186		

[illegible]

RESULT 13

US-08-577-483-14/c
? Sequence 14, Application US/08577483
? Patent No. 6100451
? GENERAL INFORMATION:
? APPLICANT: Chappell, Joseph
? APPLICANT: Yin, Staohui
? APPLICANT: Cornett, Catherine A.G.
? TITLE OF INVENTION: Transcriptional Control Sequences and
? TITLE OF INVENTION: Methods
? NUMBER OF SEQUENCES: 14
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson, P.C.
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: MA
? COUNTRY: US
? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/577,483
? FILING DATE: 22-DEC-1995
? CLASSIFICATION: 80C
? ATTORNEY/AGENT INFORMATION:
? NAME: Paul T. Clark
? REGISTRATION NUMBER: 32,164
? REFERENCE/DOCKET NUMBER: 07678/003001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 542-5070
? TELEFAX: (617) 542-8906
? TELEX: 200154
? INFORMATION FOR SEQ ID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1368 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: both
? MOLECULE TYPE: cdna
? (US-08-577-483-14)

[illegible]

	Matches	182;	Conservative	0;	Mismatches	188;	Indels	6;	Gaps	21
QY	90	TGTCAGTATGTGGATTATCTTGTGAAAGACAGGAGTGAAGAACATTTTGTGTAATGG	149							
Db	4277	TGTAAATCTCGTACAATATTTCAATGACAAAGGTGTAAAGGTATCTATGTAATGG	4218							
QY	150	CACACAGAGAGAGGCTGTCCCTGAGCGCTTCAGAGCGTCGCCAGGTTGCGAGAGAGTG	209							
Db	4217	TTCTTCAGGTGAATGTATTACCAAAAGTGTAGAAGATCGTAACAAATTTATTGAGCTGT	4158							
QY	210	GGTGACAAAAGGGAAGACAAGCTGGATCAGGTGTGAATTCAGCTAGGAGCACTGAGCTT	269							
Db	4157	TATGGAGTTGCTRAAGGTAAATTT--AACAGTTATCAACCATATTCGATGTAAATAACAC	4101							
QY	270	GAAGGAGTCAAGGAACTGGCCCCAACATGACGACAGAAATAGGAGCTGATGGCATCTGCT	329							
Db	4100	GAAGAGTAGTATCGAATTTGCAAAACATTCAGAAAGTTGTGGAGTCGATCTATTGCAGC	4041							
QY	330	CATTGCAAGCTTCTTCCTGAAGCCATGGACCAAGATATCCTCATTAATTTCTCTAAGGA	389							
Db	4040	TATCCCAACCTATTATTATTTCAA---ATTGCCAGAGTATCAATCGCAGCATATTGGATGC	3984							
QY	390	AGTGGCTGTGCGCGCCCTCGCCCTGTCATTTATTTACTATCACATTCCTCCCTTGACAGG	449							
Db	3983	AATGAGTGAAGCTCGGTCAATACAGATTTTATTATCTATAATATTCACAAATTGGCAGG	3924							
QY	450	GGTAAGAATTCTGTGCT	465							
Db	3923	GGTTGGCTTGACTGGT	3908							

```

1  RESULT 13
2  US-08-443-639-7/c
3  Sequence 7, Application US/08443639
4  Patent No. 5981843
5  GENERAL INFORMATION:
6  APPLICANT: Chappell, Joseph
7  APPLICANT: Yin, Shaohui
8  APPLICANT: Cornett, Catherine A.G.
9  TITLE OF INVENTION: Transcriptional Control Sequences and
10  TITLE OF INVENTION: Methods
11  NUMBER OF SEQUENCES: 13
12  CORRESPONDENCE ADDRESS:
13  ADDRESSEE: Greenlee and Winner, P.C.
14  STREET: 5370 Manhattan Circle, Suite 201
15  CITY: Boulder
16  STATE: CO
17  COUNTRY: US
18  ZIP: 80303
19  COMPUTER READABLE FORM:
20  MEDIUM TYPE: Floppy disk
21  COMPUTER: IBM PC compatible
22  OPERATING SYSTEM: PC-DOS/MS-DOS
23  SOFTWARE: PatentIn Release #1.0, Version #1.30
24  CURRENT APPLICATION DATA:
25  APPLICATION NUMBER: US/08/443,639
26  FILING DATE: 18-MAY-1995
27  CLASSIFICATION: 800
28  ATTORNEY/AGENT INFORMATION:
29  NAME: Ferber, Donna M.
30  REGISTRATION NUMBER: 33878
31  REFERENCE/DOCKET NUMBER: 69-94
32  TELECOMMUNICATION INFORMATION:
33  TELEPHONE: (303) 493-8860
34  TELEFAX: (303) 499-8089
35  INFORMATION FOR SEQ ID NO: 7:
36  SEQUENCE CHARACTERISTICS:
37  LENGTH: 4254 base pairs
38  TYPE: nucleic acid
39  STRANDEDNESS: double
40  TOPOLOGY: linear
41  MOLECULE TYPE: DNA (genomic)
42  HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana tabacum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1217...1327, 1455..1718, 1806..2182,
; LOCATION: 2259
; LOCATION: ..2477, 2609..2747, 2903..3148, 3262..3558)
; US-08-443-639-7

Query Match      2.9%; Score 41.8; DB 2; Length 4254;
Best Local Similarity 50.2%; Pred. No. 3.047;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY      1091  CACAGATTTTGTGGAGAAATTCGTTTATCGATGAATGGATCAAGAGAAA 1150
DB      353  CACACATACATTATATATATATATATATATATATATATATATATATATATA 294

QY      1151  TTCTAATTCGATTAATCCCATCTGCTTTAGGAGCTCTCATTATCGGTCTCTGGTTCCT 1210
DB      293  TATATATATTAATTCCTAATTTTCCTTTTAACTTTTCACAATTTAGTGTATTTTCTG 234

QY      1211  AATCCCTATTTTAAAGTTGTCCTAATTTTAAACCACTATAPATGTCCTCATTTAATAAAT 1270
DB      233  ATTATATAATGATTTTATAAAATTTATAAATATAAAATAAAACACCATGGGCTTACA 174

QY      1271  ATTCATTGGAAATCTAGGAAAATCT 1295
DB      173  CTTCTGCTTCGTGTAGATAAAGCG 149

```

RESULT 14
US-09-435-380-7/c
Sequence 7, Application US/09435380
Patent No. 6605764
GENERAL INFORMATION:
APPLICANT: Chappelli, Joseph
APPLICANT: Yin, Shaohui
APPLICANT: Cornett, Catherine A.G.
TITLE OF INVENTION: Transcriptional Control Sequences and
TITLE OF INVENTION: Methods
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/435,380
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/443,639
FILING DATE: 18-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 69-94
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4254 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

```
;
;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana tabacum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1217..1327, 1455..1718, 1836..2182,
; LOCATION: 2259
; LOCATION: ..2477, 2609..2747, 2903..3148, 3262..3558);
US-89-435-380-7

Query Match          2.9%; Score 41.8; DB 4; Length 4254;
Best Local Similarity 50.2%; Pred. No. 0.047;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1091 CACGATTTTTTTGTGGAGAAATTCCTGTTTATATGGATGAAATCAAGAGGAAA 1150
DB 353 CACACATACATTTATATATATATATATATATATATATATATATATATATATAT 294
QY 1151 TTGTAATGANTTAATCCATCTGCTTTAGGAGCTCTCATTAATCTCGGTCCTCT 1210
DB 293 TATATACATATTAATTCATATATTTCTTTAACTTTTCACAATTTTAGTGTATTT 234
QY 1211 AATCCTATTTTAAAGTCTGCTAAATTTTAAACCACTATATATATGCTTCTCATTTTATATAAT 1272
DB 233 ATTATATATATGATTTTATATATTTTATATATATATATATATATATATATATAT 174
QY 1271 ATTCAATTTGGAACTTAGGAAACTC 1295
DB 173 CCTCGCTTCGTGTAGATAAAAGCG 143

RESULT 15
US-08-956-171E-38/c
; Sequence 38, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; GIL H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannin
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknwn>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
```

```
;
;
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23439 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-08-956-171E-38

Query Match          2.9%; Score 41.8; DB 4; Length 23439;
Best Local Similarity 47.4%; Pred. No. 0.12;
Matches 158; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 13 AAGAGAAACCTTCAGGGTCTTGTGGCTGCAACCATCAAGCCCAATGACTCAGAAATGAGAA 72
DB 8682 AACAAAGATTTTAAAGGTTTATATGACGGTTACTCTGTTCTTTTCATGAAATGCTCAA 8623
QY 73 ATCAACTTTTCGTAATTTGGTCAGTATGTTGGATTATCTTGTGAAGAACAGGGAGTGAAG 132
DB 8622 GTAAATGAACAAGGACITTAACAAATTTCTCAAAATGCCAATGMAAATGAAATTAGAC 8563
QY 133 AACATTTTGTGAATGGCACACACAGGAGAGGCCCTGTCCCTGAGCGTCTCAGAGCGTGC 192
DB 8562 GGTCTTTATGTAATGTTAGTCTCGGCCGAAACTTTTTTATTAA---ATACAGAGCAGAA 8506
QY 193 CAGGTTGCAGAGGAGTGGGTGACAAAGGGAAGCAAGCTGGATCAGGTGATAATTCAC 252
DB 8505 AAGCAAGTTTTCAGGTTGCCAAAGAGGCGAGTTGGGGATAAAGTGAAATTTGATTGCTCAA 8446
QY 253 GTAGGAGCACTCAGCTTGAAGGAGTCAAGGAACTGGCCCACTGCAGCAGCAAAATGGA 312
DB 8445 GTAGTTGCTTAGATTAAATGAAGCCATTGAAGTTGGAAATATATGCCACAGAACTCGGT 8386
QY 313 GCTGATGGCATCGGTGTCTCATTTGCACCGTTCTTC 345
DB 8385 TATGATGCACCTTTCGCGGTAAACCATTTCTAC 8353
```

Search completed: May 11, 2004, 07:22:11
Job time : 127.968 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 11:46:35 : Search time 662.155 Seconds
(without alignments)
9776.195 Million cell updates/sec

Title: US-09-930-440B-1

Perfect score: 1429

Sequence: 1 atggccttcaccaagaagaa.....ctttgaattcatttcgatg 1429

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 588172

Minimum DB seq length: 5

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1429	100.0	1429	9	US-09-984-205-5
2	1429	100.0	1429	9	US-09-930-440B-1
3	1326	92.9	1552	15	US-10-102-524-1751
4	792	55.4	1923	9	US-09-833-381-1100
5	654	45.8	1127	9	US-09-833-381-1099
6	374	26.2	374	15	US-10-102-524-2020
7	312	21.8	400	13	US-10-085-783A-21707
8	312	21.8	400	16	US-10-242-535A-21707
9	149.2	10.4	309	9	US-09-783-590-1834
10	128.2	9.0	472	10	US-09-918-995-8879
11	46.4	3.2	177249	16	US-10-085-117-223
12	45.4	3.2	6117	15	US-10-311-455-998
13	44.8	3.1	5771	15	US-10-311-455-1924
14	44.4	3.1	6109	13	US-10-221-613-33

15	44.4	3.1	6109	15	US-10-311-455-299	Sequence 289, App
16	44.2	3.1	19380	13	US-10-221-613-389	Sequence 389, App
17	44	3.1	8170	15	US-10-240-453-131	Sequence 131, App
18	44	3.1	13574	15	US-10-311-455-1290	Sequence 1290, App
19	43.2	3.0	6361	15	US-10-311-455-1114	Sequence 1114, App
20	43.2	3.0	9897	13	US-10-158-844-10	Sequence 10, App
21	43.2	3.0	17848	15	US-10-239-676-28	Sequence 28, App
22	43.2	3.0	17848	15	US-10-240-453-38	Sequence 38, App
23	43.2	3.0	17848	17	US-10-257-166-58	Sequence 58, App
24	42.2	3.0	8991	13	US-10-221-613-256	Sequence 256, App
25	42	2.9	6185	15	US-10-311-455-385	Sequence 385, App
26	42	2.9	3673778	15	US-10-312-841-2	Sequence 2, App
27	41.8	2.9	5185	15	US-10-311-455-1007	Sequence 1007, App
28	41.8	2.9	10945	15	US-10-240-453-227	Sequence 227, App
29	41.8	2.9	23439	8	US-08-781-386A-38	Sequence 38, App
30	41.8	2.9	23439	13	US-10-329-624-38	Sequence 38934, A
31	41.6	2.9	490	13	US-10-027-632-38934	Sequence 83646, A
32	41.6	2.9	490	13	US-10-027-632-83646	Sequence 38934, A
33	41.6	2.9	490	16	US-10-027-632-38934	Sequence 83646, A
34	41.6	2.9	490	16	US-10-027-632-83646	Sequence 126910, A
35	41.6	2.9	800	13	US-10-027-632-126910	Sequence 126910, A
36	41.6	2.9	800	16	US-10-027-632-126910	Sequence 45335, A
37	41.6	2.9	885	16	US-10-369-493-45335	Sequence 202, App
38	41.6	2.9	6049	15	US-10-311-455-202	Sequence 2154, App
39	41.6	2.9	7623	15	US-10-311-455-2154	Sequence 28, App
40	41.4	2.9	2637	15	US-10-176-306-28	Sequence 32, App
41	41	2.9	8323	15	US-10-311-455-32	Sequence 1686, App
42	41	2.9	13326	15	US-10-311-455-1686	Sequence 45, App
43	41	2.9	15479	17	US-10-257-166-45	Sequence 1251, App
44	40.8	2.9	7255	15	US-10-311-455-1251	Sequence 1084, App
45	40.8	2.9	10957	15	US-10-311-455-1084	

ALIGNMENTS

RESULT 1
US-09-984-205-5
Sequence 5, Application US/09984205
Patent No. US20020137175A1
GENERAL INFORMATION:
APPLICANT: Coleman, Timothy A. et al.
TITLE OF INVENTION: Human Glycosylation Enzymes
FILE REFERENCE: EP0505D1
CURRENT APPLICATION NUMBER: US/09/984,205
CURRENT FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: PCT/US00/05325
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 03/516,143
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/122,409
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1429
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(590)
US-09-984-205-5

Query Match 100.0%; Score 1429; DB 9; Length 1429;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGCCTTCACCAAGAAAGAACTTCAGGCTTCCTGCTGCACCAATCAGCCCAATGACT	60
Db	1	ATGGCCTTCACCAAGAAAGAACTTCAGGCTTCCTGCTGCACCAATCAGCCCAATGACT	60
QY	61	GAGAAATGGAAATCACTTTTCAGTAATGGTCAGTATGTTGTAATTCCTTGTGAAGAA	120

Db 61 GAGAAATGAGAAATCAACTTTTCAGTAATTTGGTCGATATGGAATATCTTGTGAAGAA 120
 QY 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAACAGGAGAGGAGGCTGTCCCTGAGGGTC 180
 Db 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAACAGGAGAGGAGGCTGTCCCTGAGGGTC 180
 QY 181 TCAGAGCGTGGCCAGGTTGACAGAGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 240
 Db 181 TCAGAGCGTGGCCAGGTTGACAGAGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 240
 QY 241 GTGATTAATTCAGTGAAGAGCACTCAGCTTGAAGAGTCAACAGGAGTGGCCCAACATGCA 300
 Db 241 GTGATTAATTCAGTGAAGAGCACTCAGCTTGAAGAGTCAACAGGAGTGGCCCAACATGCA 300
 QY 301 GCAGAAATAGGAGTGTGATGAGTGGTGTCAATTCACACCGTTCTTCTTCAAGCCATGAC 360
 Db 301 GCAGAAATAGGAGTGTGATGAGTGGTGTCAATTCACACCGTTCTTCTTCAAGCCATGAC 360
 QY 361 AAGATATCTGATTAATTTCTTAAGGAGTGGTGTGCTGCGCCGCTGCTGCTGCTGCTGCT 420
 Db 361 AAGATATCTGATTAATTTCTTAAGGAGTGGTGTGCTGCGCCGCTGCTGCTGCTGCTGCT 420
 QY 421 TATTACTATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 Db 421 TATTACTATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 QY 481 GGGATTCGGTAAGATCCCACTTCCAGGGTGAATTCAGTGAATTCAGTGAATTCAGTGA 540
 Db 481 GGGATTCGGTAAGATCCCACTTCCAGGGTGAATTCAGTGAATTCAGTGAATTCAGTGA 540
 QY 541 GACTTCGGGCAATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 Db 541 GACTTCGGGCAATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 QY 601 GATGAGCAACTGTTGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 660
 Db 601 GATGAGCAACTGTTGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 660
 QY 661 TCCAGAGATTTATCAACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 720
 Db 661 TCCAGAGATTTATCAACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 720
 QY 721 TCATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 Db 721 TCATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 QY 781 CCAGGAGTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Db 781 CCAGGAGTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 QY 841 CTGATTTAAGGATGGAACCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 900
 Db 841 CTGATTTAAGGATGGAACCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 900
 QY 901 TTGACCTTTGAGACATAATCTACCTTAATAGTGAATTTTTCAGGGAATTTAGAT 960
 Db 901 TTGACCTTTGAGACATAATCTACCTTAATAGTGAATTTTTCAGGGAATTTAGAT 960
 QY 961 GAACTTGAATAAATCTCTAGCAATGAATTCACAAATAGCAATGAGTGAATTTG 1020
 Db 961 GAACTTGAATAAATCTCTAGCAATGAATTCACAAATAGCAATGAGTGAATTTG 1020
 QY 1021 TGAGCCTTAAAGTCTTATTTTGTGAAGGAGGCAAAACTCTAGGAGTCAACACTCTCAG 1080
 Db 1021 TGAGCCTTAAAGTCTTATTTTGTGAAGGAGGCAAAACTCTAGGAGTCAACACTCTCAG 1080
 QY 1081 TCATTCATTTACAGATTTTGTGGAATTTTCTGTTATATGATGATGATGATGATGATGAT 1140
 Db 1081 TCATTCATTTACAGATTTTGTGGAATTTTCTGTTATATGATGATGATGATGATGATGAT 1140
 QY 1141 AAGAGGAAATTTGATTAATTTTCAATCTCTCTTTAGGAGCTCTCAATATCTCGGTC 1200
 Db 1141 AAGAGGAAATTTGATTAATTTTCAATCTCTCTTTAGGAGCTCTCAATATCTCGGTC 1200

RESULT 2

US-09-930-440B-1
 ; Sequence 1, Application US/09930440B
 ; Patent No. US20020142386A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Betenbaugh et al.
 ; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
 ; FILE REFERENCE: PF50392
 ; CURRENT APPLICATION NUMBER: US/09/930,440B
 ; CURRENT FILING DATE: 2001-08-16
 ; PRIOR APPLICATION NUMBER: 60/227,579
 ; PRIOR FILING DATE: 2000-08-25
 ; PRIOR APPLICATION NUMBER: 09/516,793
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/169,624
 ; PRIOR FILING DATE: 1999-12-08
 ; PRIOR APPLICATION NUMBER: 60/122,582
 ; PRIOR FILING DATE: 1999-03-02
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 1429
 ; TYPE: DNA
 ; ORGANISM: Hmo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(693)
 US-09-930-440B-1

Query Match 100.0%; Score 1429; DB 9; Length 1429;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCCTCCAAAGAGAACTTCAGGCTTGTGGCTGCAACCATCAGCCCAATGACT 60
 Db 1 ATGGCCCTCCAAAGAGAACTTCAGGCTTGTGGCTGCAACCATCAGCCCAATGACT 60
 QY 61 GAGATGAGAGAAATCAACTTTTCAGTAAATTCGTTCAGTATGATGATTTCTTGTGAAGAA 120
 Db 61 GAGATGAGAGAAATCAACTTTTCAGTAAATTCGTTCAGTATGATGATTTCTTGTGAAGAA 120
 QY 121 CAGGAGTGAAGAAATTTTGTGAATGGCAACACAGGAGAGGCTGTCCCTGAGGCTC 180
 Db 121 CAGGAGTGAAGAAATTTTGTGAATGGCAACACAGGAGAGGCTGTCCCTGAGGCTC 180
 QY 181 TCAGAGCTCCCGAGTTCAGAGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 240
 Db 181 TCAGAGCTCCCGAGTTCAGAGAGTGGGTGACAAAGGGAAGGACAGCTGGATCAG 240
 QY 241 GTGATTAATTCAGTGAAGAGCACTCAGCTTGAAGAGTCAACAGGAGTGGCCCAACATGCA 300
 Db 241 GTGATTAATTCAGTGAAGAGCACTCAGCTTGAAGAGTCAACAGGAGTGGCCCAACATGCA 300
 QY 301 GCAGAAATAGGAGTGTGATGAGTGGTGTCAATTCACACCGTTCTTCTTCAAGCCATGAC 360

301	GCAGAAATAGGAGCTGATGGCATCGCTGCTCATTTGCACCGGTTCTTCTCTCAAGCCATGGACC	360
361	AAAGATATCTCGATTAAATTTTCTTAAAGGAAGTGGCTGCTCGCGCCGCTGCGCTGCCATTT	420
361	AAAGATATCTCGATTAAATTTTCTTAAAGGAAGTGGCTGCTCGCGCCGCTGCGCTGCCATTT	420
421	TATTACTATCACAAATTCCTGCTTGACACGGGGTAAAGATTGCTGCTGAGAGATGTTGGGAT	480
421	TATTACTATCACAAATTCCTGCTTGACACGGGGTAAAGATTGCTGCTGAGAGATGTTGGGAT	480
481	GGGATTTCTGATAGATGCCACCTTCCAAAGGGCTGAATTCAGTGATACAGATCTCTTA	540
481	GGGATTTCTGATAGATGCCACCTTCCAAAGGGCTGAATTCAGTGATACAGATCTCTTA	540
541	GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTGCCTTTTGGGGTG	600
541	GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAACAGTTTGCCTTTTGGGGTG	600
601	GATGAGCAACTGTTGATGCTCTGTTGATGGGAGCAACTGGAGCAGTGGGACGTTTGTGA	660
601	GATGAGCAACTGTTGATGCTCTGTTGATGGGAGCAACTGGAGCAGTGGGACGTTTGTGA	660
661	TCCAGAGATTATCAACTTTGTTGTCAAACTAGCTTTTGGAGTGTCACACCAAGCCCA	720
661	TCCAGAGATTATCAACTTTGTTGTCAAACTAGCTTTTGGAGTGTCACACCAAGCCCA	720
721	TCATGACTCTGGTCTCTGGGATTCCAATGGGCCCAACCCCGGCTTCCACTGCAGAAAGCCT	780
721	TCATGACTCTGGTCTCTGGGATTCCAATGGGCCCAACCCCGGCTTCCACTGCAGAAAGCCT	780
781	CCAGGGAGTTTACTGATAGTCTGAAGCTATAAATCTGAAGAGCCTGGATTTCCCTTCTTTCA	840
781	CCAGGGAGTTTACTGATAGTCTGAAGCTATAAATCTGAAGAGCCTGGATTTCCCTTCTTTCA	840
841	CTGATTTAAAGGATGGAACTGCGGAAGCTGATAGTGCCTCTCTATCAATCAGGGT	900
841	CTGATTTAAAGGATGGAACTGCGGAAGCTGATAGTGCCTCTCTATCAATCAGGGT	900
901	TTGCACCTTGACACATAATCTACCTTAAATAGTGCAATTTTCTCAGGGAAATTTTAGAT	960
901	TTGCACCTTGACACATAATCTACCTTAAATAGTGCAATTTTCTCAGGGAAATTTTAGAT	960
961	GAACTTGAAATAACTCTCTAGCAAAATGAAATCTCAAAATGAGCAATGAGGTACCTTTTG	1020
961	GAACTTGAAATAACTCTCTAGCAAAATGAAATCTCAAAATGAGCAATGAGGTACCTTTTG	1020
1021	TSAGCCCTTAAAGAGTCTTATTTTGTGAAGGGGCAAAACCTCTAGGAGTCACACTCTCAG	1080
1021	TSAGCCCTTAAAGAGTCTTATTTTGTGAAGGGGCAAAACCTCTAGGAGTCACACTCTCAG	1080
1081	TCATTCATTTACAGATTTTTTGTGAGGAAATTTCTGTTTATATGATGAATGGAATC	1140
1081	TCATTCATTTACAGATTTTTTGTGAGGAAATTTCTGTTTATATGATGAATGGAATC	1140
1141	AAGAGGAAATTTGTAATTTGATTAATTCATCTCTGCTTTAGGAGCTCTCATTTCTGGTC	1200
1141	AAGAGGAAATTTGTAATTTGATTAATTCATCTCTGCTTTAGGAGCTCTCATTTCTGGTC	1200
1201	TCGTTTCTTAATCCTATTTTAAAGTGTCTTAATTTTAAACCACTATAATGCTTCAT	1260
1201	TCGTTTCTTAATCCTATTTTAAAGTGTCTTAATTTTAAACCACTATAATGCTTCAT	1260
1261	TTTAAATAAATTTCAATTTGGAATCTAGAAAACTCTGAGCTACTGCAATTTAGGACGAC	1320
1261	TTTAAATAAATTTCAATTTGGAATCTAGAAAACTCTGAGCTACTGCAATTTAGGACGAC	1320
1321	TTTAAATACCAACTGTAACTATGCTCAACTGTATACAACCTCAAAATACACCGCTCATTT	1380
1321	TTTAAATACCAACTGTAACTATGCTCAACTGTATACAACCTCAAAATACACCGCTCATTT	1380
1381	GGCTGCTCAGTCTAACTCTAGAAATGGAGCTTTTGGAAATTCATTTGGATG	1423
1381	GGCTGCTCAGTCTAACTCTAGAAATGGAGCTTTTGGAAATTCATTTGGATG	1423

```

RESULT 3
US-10-102-524-1751
; Sequence 1751, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT OF KIDNEY CANCER
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1751
; LENGTH: 1552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-1751

Query Match          92.94;   Score 1328;   DB 15;   Length
Best Local Similarity 94.44;   Fred. No. 0;
Matches 1423;   Conservative 0;   Mismatches 0;   Indels

```

RESULT 3

```
Qy 651 ----- 650
Db 705 AACTACCTGGGAAAAAGACAAACACAGATGTTGGAGGCTTTTGAACAAAGACTTCCT 764
Qy 651 ----- CAGTTTGTATCCAGAGATTATCAACTTGTGTCAAACTAGGT 695
Db 765 TTAGCCCTGAACTATCACTGTTTGTATCCAGAGATTATCAACTTGTGTCAAACTAGGT 824
Qy 696 TTGGAGTGTACAGACCAAGACCAATCATATGATCTGTGCTCTGGATTCGAATCCAAATGGGCCCA 755
Db 825 TTGGAGTGTACAGACCAAGACCAATCATATGATCTGTGCTCTGGATTCGAATCCAAATGGGCCCA 884
Qy 756 CCCGGCTTCCAGTGCAGAAAGCCCTCCAGGGAGTTTACTGATAGTGCCTGGAAGCTAAATCTG 815
Db 885 CCCGGCTTCCAGTGCAGAAAGCCCTCCAGGGAGTTTACTGATAGTGCCTGGAAGCTAAATCTG 944
Qy 816 AAGAGCTGTGATTTCCCTTTCTTTCACTGATTTAAAGGATGAAACTTGGAAAGCTGGTAGC 875
Db 945 AAGAGCTGTGATTTCCCTTTCTTTCACTGATTTAAAGGATGAAACTTGGAAAGCTGGTAGC 1004
Qy 876 TAGTGCCTCTCTATCAATCAGGCTTTCACCTTGGAGACATTAATCTACCTTAATAGTAGC 935
Db 1005 TAGTGCCTCTCTATCAATCAGGCTTTCACCTTGGAGACATTAATCTACCTTAATAGTAGC 1064
Qy 936 ATTTTCTTCCAGGATTTTGTAGTGAATCTTGAATAAACTCTCTAGCAAAATGAAATCTC 995
Db 1065 ATTTTCTTCCAGGATTTTGTAGTGAATCTTGAATAAACTCTCTAGCAAAATGAAATCTC 1124
Qy 996 ACAATAAGCAATGAGGTACCTTTTGTAGCCCTTAAAAAGTCTTATTTTGTGAAGGGGCAA 1055
Db 1125 ACAATAAGCAATGAGGTACCTTTTGTAGCCCTTAAAAAGTCTTATTTTGTGAAGGGGCAA 1184
Qy 1056 AAATCTAGGAGTACAACTCTCAGTCACTCATTTCAAGATTTTGTGAGAGAAATTT 1115
Db 1185 AAATCTAGGAGTACAACTCTCAGTCACTCATTTCAAGATTTTGTGAGAGAAATTT 1244
Qy 1116 CTGTTTATATGATGAAATGGAATCAAGAGAAATTTGAATTAATTAATTCATCTCTC 1175
Db 1245 CTGTTTATATGATGAAATGGAATCAAGAGAAATTTGAATTAATTAATTCATCTCTC 1304
Qy 1176 TTTAGGAGCTCTCAATTAATCTCGGCTCTGTTTCTTAATTCCTTAATTTAAAGTTTCTAAT 1235
Db 1305 TTTAGGAGCTCTCAATTAATCTCGGCTCTGTTTCTTAATTCCTTAATTTAAAGTTTCTAAT 1364
Qy 1236 TTAACCACTAATATATGCTCTCACTTTAATAATATTCATTTGGAATCTAGAAACTC 1295
Db 1365 TTAACCACTAATATATGCTCTCACTTTAATAATATTCATTTGGAATCTAGAAACTC 1424
Qy 1296 TGAGTACTGCAATTTAGGAGGCACTTTAATACCAAACTGTAACATGCTCAACTGTATA 1355
Db 1425 TGAGTACTGCAATTTAGGAGGCACTTTAATACCAAACTGTAACATGCTCAACTGTATA 1484
Qy 1356 CAATCAAAATACACAGCTCATTTGGTGTCTAGTCTAATCTTAGAATGGAATGCTTTTG 1415
Db 1485 CAATCAAAATACACAGCTCATTTGGTGTCTAGTCTAATCTTAGAATGGAATGCTTTTG 1544
Qy 1416 AATTCATT 1423
Db 1545 AATTCATT 1552
```

RESULT 4
US-09-930-381-1100/c
; Sequence 1100, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5820-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29

```
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1150  
; LENGTH: 1923  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-930-381-1100
```

Query Match 55.4%; Score 792; DB 9; Length 1923;
Best Local Similarity 91.2%; Pred. No. 2.1e-217;
Matches 899; Conservative 0; Mismatches 0; Indels 87; Gaps 2;

```
Qy 531 AGATCTCTTAGACTTCGGCAATGTGTGATCAGCAATCCAGCAACAGTGTTCCTTCTCT 590
Db 1923 AGATCTCTTAGACTTCGGCAATGTGTGATCAGCAATCCAGCAACAGTGTTCCTTCTCT 1864
Qy 591 TTTTGGGTGGATGAGCAACTGTGTGAGTCTCTGCTGTGTGATGGAGCAACTGGAGCAGTGG 650
Db 1863 TTTTGGGTGGATGAGCAACTGTGTGAGTCTCTGCTGTGTGATGGAGCAACTGGAGCAGTGG 1804
Qy 651 ----- 650
Db 1803 CAGTACCTATAACTACCTGGGAAAAAGACAAACACAGATGTTGGGGGCTTTTGAACRAAA 1744
Qy 651 ----- CAGTTTGTATCCAGAGATTATCAACTTTGTG 684
Db 1743 GGGACTTCTCTTTAGCCCTGAACTATCATGTTTGTATCCAGATTTTCAACTTTGTG 1684
Qy 685 TCAAACTAGGTTTGGAGTGTACAGACCAAAGCCATCATGCTCTGCTCTGGGATTC 744
Db 1683 TCAAACTAGGTTTGGAGTGTACAGACCAAAGCCATCATGCTCTGCTCTGGGATTC 1624
Qy 745 CAATGGGGCCACCCCGGCTTCCAGTGCAGAAAGCCTCCAGGAGTTTACTGATAGTCTG 804
Db 1623 CAATGGGGCCACCCCGGCTTCCAGTGCAGAAAGCCTCCAGGAGTTTACTGATAGTCTG 1564
Qy 805 AAGCTAAACTCAAGAGCCTGGATTTCTTCTTCACTGATTTAAAGGATGAAACTTGG 864
Db 1563 AAGCTAAACTCAAGAGCCTGGATTTCTTCTTCACTGATTTAAAGGATGAAACTTGG 1504
Qy 865 AAGCTGTGTAGTGTGCTCTCTATCAAAATCAGGTTTGGACCTTGGAGCAATATCTACC 924
Db 1503 AAGCTGTGTAGTGTGCTCTCTATCAAAATCAGGTTTGGACCTTGGAGCAATATCTACC 1444
Qy 925 TTAATAGTGCATTTTCTCAGGAAATTTAGATGAATTAATTAATTAATTAATTAATTA 984
Db 1443 TTAATAGTGCATTTTCTCAGGAAATTTAGATGAATTAATTAATTAATTAATTAATTA 1384
Qy 985 AATGAATCTCACAATAGCAATGAGGTACCTTTTGTGAGCCTTAAAAAGTCTTATTTTG 1044
Db 1383 AATGAATCTCACAATAGCAATGAGGTACCTTTTGTGAGCCTTAAAAAGTCTTATTTTG 1324
Qy 1045 TGAAGGGCAAAAACTCTAGAGTCAAACTCTCAGTCAATTCATTTCTCAGATTTTGTG 1104
Db 1323 TGAAGGGCAAAAACTCTAGAGTCAAACTCTCAGTCAATTCATTTCTCAGATTTTGTG 1264
Qy 1105 TGGAGAAATTTCTCTTATATGATGAAATGGAATCAAGGAAATTTGTAATTCATTAA 1164
Db 1263 TGGAGAAATTTCTCTTATATGATGAAATGGAATCAAGGAAATTTGTAATTCATTAA 1204
Qy 1165 TTCCATCTCTTTAGGAGCTCTCATTTATCTCGGCTCTGCTGCTCTTAATCTTATTTAAA 1224
Db 1203 TTCCATCTCTTTAGGAGCTCTCATTTATCTCGGCTCTGCTGCTCTTAATCTTATTTAAA 1144
Qy 1225 GTTCTCTAATTTTAAACCACTATATATGCTCTCATTTTAAATAATATTCATTTGGAATC 1284
Db 1143 GTTCTCTAATTTTAAACCACTATATATGCTCTCATTTTAAATAATATTCATTTGGAATC 1084
Qy 1285 TAGGAATCTCTGAGTACTGCAATTTAGGAGGCACTTTTAAATCCAACTGTACATCTC 1344
Db 1083 TAGGAATCTCTGAGTACTGCAATTTAGGAGGCACTTTTAAATCCAACTGTACATCTC 1024
Qy 1345 TCAACTG-NTACAACTCAAAATACACCACTCAATTTGGCTGCTCAGTCTAATCTCTAGAAT 1404
```

Db 1023 TCAACTGTATCAACTCAAAATACACAGCTCATTTGGCTGCTCAGTCTAACTCTAGAAAT 964
 QY 1405 -GGATGCTTTGAATTCATTTCCGATG 1429
 Db 963 GGGAGCTTTTGAATTCATTTCCGATG 938

RESULT 5
 US-09-833-381-1099
 ; Sequence 1099, Application US/09833381
 ; Patent No. US20020132090A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
 ; FILE REFERENCE: 5800-119
 ; CURRENT APPLICATION NUMBER: US/09/833,381
 ; CURRENT FILING DATE: 2001-04-11
 ; PRIOR APPLICATION NUMBER: 09/516,448
 ; PRIOR FILING DATE: 2000-02-29
 ; NUMBER OF SEQ ID NOS: 2050
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1099
 ; LENGTH: 1127
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-833-381-1099

Query Match 45.8%; Score 654; DB 9; Length 1127;
 Best Local Similarity 100.0%; Pred. No. 1.1e-177; Mismatches 0; Indels 0; Gaps 0;
 Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGGGTGCAACCATCACGCGCAATGACT 60
 Db 199 ATGGCTTCCCAAGAGAACTTCAGGCTCTTGGGTGCAACCATCACGCGCAATGACT 258
 QY 61 GAGATGAGAAATCAACTTTTCATATTTGTCAGTATGTCAGTATTTGTCAGAA 120
 Db 259 GAGATGAGAAATCAACTTTTCATATTTGTCAGTATGTCAGTATTTGTCAGAA 318
 QY 121 CAGGAGTGAAGAACTTTTGTGATGCAACAGAGAGAGGCTGTCTCCTGAGCGTC 180
 Db 319 CAGGAGTGAAGAACTTTTGTGATGCAACAGAGAGAGGCTGTCTCCTGAGCGTC 378
 QY 181 TCAGAGCTGCGCAGGTTGCAGAGAGTGGGTGACAAAAGGAGGAGCAAGCTGGATCAG 240
 Db 379 TCAGAGCTGCGCAGGTTGCAGAGAGTGGGTGACAAAAGGAGGAGCAAGCTGGATCAG 438
 QY 241 GTGATTAATTCAGTGAAGCACTGAGCTTGAAGAGTCAAGAACTGGCCCAACATGCA 300
 Db 439 GTGATTAATTCAGTGAAGCACTGAGCTTGAAGAGTCAAGAACTGGCCCAACATGCA 498
 QY 301 CGAGAAATAGAGCTGATGGCATCCTGTCTATTGCAACGCTTCTCTCAAGCCATGGACC 360
 Db 499 CGAGAAATAGAGCTGATGGCATCCTGTCTATTGCAACGCTTCTCTCAAGCCATGGACC 558
 QY 361 AAGATATCTCTGATTAATTTCTTAAGGAAAGTGGCTGTGCGCCGCTGCGCATTT 420
 Db 559 AAGATATCTCTGATTAATTTCTTAAGGAAAGTGGCTGTGCGCCGCTGCGCATTT 618
 QY 421 TATTACTATCACTTCTGCTTGAAGAGTCAAGGCTTGAAGAGTCAAGGAGTGTGGAT 480
 Db 619 TATTACTATCACTTCTGCTTGAAGAGTCAAGGCTTGAAGAGTCAAGGAGTGTGGAT 678
 QY 481 GGGATTCCTGGATGAATCCCACTTCCAAAGGCTGAAATTCAGTCAATCATCTCTTA 540
 Db 679 GGGATTCCTGGATGAATCCCACTTCCAAAGGCTGAAATTCAGTCAATCATCTCTTA 738
 QY 541 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAAGTTCCTTTTCTTTTGGGGTG 600
 Db 739 GACTTCGGGCAATGTGTTGATCAGAAATCGCCAGCAAGTTCCTTTTCTTTTGGGGTG 798
 QY 601 GATGAGCAACTGTTGAGTGTCTGTGATGGAGCAACTGAGAGCAGTGGGCAGT 654

Db 799 GATGAGCAACTGTTGAGTGTCTGTGATGGAGCACTGGAGCAGTGGGCAGT 852

RESULT 6
 US-10-102-524-220/c
 ; Sequence 220, Application US/10102524
 ; Publication No. US20030109434A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Gaiger, Alexander
 ; APPLICANT: Gordon, Brian
 ; APPLICANT: Harlocker, Susan L.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 ; THERAPY AND DIAGNOSIS OF KIDNEY CANCER
 ; FILE REFERENCE: 210121.572
 ; CURRENT APPLICATION NUMBER: US/10/102,524
 ; CURRENT FILING DATE: 2002-03-19
 ; NUMBER OF SEQ ID NOS: 1863
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 220
 ; LENGTH: 374
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-102-524-220

Query Match 26.2%; Score 374; DB 15; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.9e-97; Mismatches 0; Indels 0; Gaps 0;
 Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 848 AAGAGATGAAACTTGGAGCTGGTAGCTAGTGTGCTCTCTATCAATCAGGGTTGCACC 907
 Db 374 AAGAGATGAAACTTGGAGCTGGTAGCTAGTGTGCTCTCTATCAATCAGGGTTGCACC 315
 QY 908 TTGAGACATTAATCTACCTTAATAGTGCATTTTTCAGGGAATTTAGTGAATTCG 967
 Db 314 TTGAGACATTAATCTACCTTAATAGTGCATTTTTCAGGGAATTTAGTGAATTCG 255
 QY 968 AATAAATCTCTAGCAATGAAATCTCAATATAGCATTTGAGTACCTTTTGTAGCCCT 1027
 Db 254 AATAAATCTCTAGCAATGAAATCTCAATATAGCATTTGAGTACCTTTTGTAGCCCT 195
 QY 1028 TAAAGATCTTATTTTGAAGAGGCAAAACTCTAGAGTCAAACTCTCAGTCTCA 1087
 Db 194 TAAAGATCTTATTTTGAAGAGGCAAAACTCTAGAGTCAAACTCTCAGTCTCA 135
 QY 1088 TTTTCAGATTTTGTGGAGAAATTTCTGTTATATGATGAAATGGAATCAAGAGGA 1147
 Db 134 TTTTCAGATTTTGTGGAGAAATTTCTGTTATATGATGAAATGGAATCAAGAGGA 75
 QY 1148 AAATTTGATTAATTCATCTCTTTAGGAGCTCTCATTTATCTCGGTCTCTGGTT 1207
 Db 74 AAATTTGATTAATTCATCTCTTTAGGAGCTCTCATTTATCTCGGTCTCTGGTT 15
 QY 1208 CCTAATCTCTATTTT 1221
 Db 14 CCTAATCTCTATTTT 1

RESULT 7
 US-10-085-783A-21707
 ; Sequence 21707, Application US/10085783A
 ; Publication No. US20040037841A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; APPLICANT: Liew, C.C.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2002
 ; CURRENT APPLICATION NUMBER: US/10/085,783A
 ; CURRENT FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13

```
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 21707
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (176)..(176)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-21707

Query Match      21.8%; Score 312; DB 13; Length 400;
Best Local Similarity 99.4%; Pred. No. 3.4e-79;
Matches 323; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGGCTTCCCAAGAGAAACTTCAGGCTTTGAGTGTGCAACCATCACGCCAATGACT 60
Db 76 ATGGCTTCCCAAGAGAAACTTCAGGCTTTGAGTGTGCAACCATCACGCCAATGACT 135
QY 61 GAGATGAGGAATCAACTTTTCAGTAATTCGTGATGATGATGATGATGATGATGAT 120
Db 136 GAGATGAGGAATCAACTTTTCAGTAATTCGTGATGATGATGATGATGATGATGAT 195
QY 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGGCTGTCCCTGAGCGTC 180
Db 196 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGGCTGTCCCTGAGCGTC 255
QY 181 TCAGAGCGTCCCGAGGTGTCAGAGAGTGGGTGACAAAAGGAGGACAAAGCTGATCAG 240
Db 256 TCAGAGCGTCCCGAGGTGTCAGAGAGTGGGTGACAAAAGGAGGACAAAGCTGATCAG 315
QY 241 GTGATAATTCAGTAGGAGCACTGAGCTTGAGGAGTGCACAGGACTGCCCAATGCA 300
Db 316 GTGATAATTCAGTAGGAGCACTGAGCTTGAGGAGTGCACAGGACTGCCCAATGCA 375
QY 301 GCA-GAAATAGGAGCTGATGGCATC 324
Db 376 GCAGGAATAGGAGCTGATGGCATC 400

RESULT 8
US-10-242-535A-21707
; Sequence 21707, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 21707
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (176)..(176)
```

```
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-21707

Query Match      21.8%; Score 312; DB 16; Length 400;
Best Local Similarity 99.4%; Pred. No. 3.4e-79;
Matches 323; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGGCTTCCCAAGAGAAACTTCAGGCTTTGAGTGTGCAACCATCACGCCAATGACT 60
Db 76 ATGGCTTCCCAAGAGAAACTTCAGGCTTTGAGTGTGCAACCATCACGCCAATGACT 135
QY 61 GAGATGAGGAATCAACTTTTTCAGTAATTCGTGATGATGATGATGATGATGATGAT 120
Db 136 GAGATGAGGAATCAACTTTTTCAGTAATTCGTGATGATGATGATGATGATGATGAT 195
QY 121 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGGCTGTCCCTGAGCGTC 180
Db 196 CAGGAGTGAAGAACATTTTGTGAATGGCAACAGGAGAGGCTGTCCCTGAGCGTC 255
QY 181 TCAGAGCGTCCCGAGGTGTCAGAGAGTGGGTGACAAAAGGAGGACAAAGCTGATCAG 240
Db 256 TCAGAGCGTCCCGAGGTGTCAGAGAGTGGGTGACAAAAGGAGGACAAAGCTGATCAG 315
QY 241 GTGATAATTCAGTAGGAGCACTGAGCTTGAGGAGTGCACAGGACTGCCCAATGCA 300
Db 316 GTGATAATTCAGTAGGAGCACTGAGCTTGAGGAGTGCACAGGACTGCCCAATGCA 375
QY 301 GCA-GAAATAGGAGCTGATGGCATC 324
Db 376 GCAGGAATAGGAGCTGATGGCATC 400

RESULT 9
US-09-783-590-1834
; Sequence 1834, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Hacdong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: P0-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1834
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (5)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (7)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (103)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (132)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (134)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

```

; LOCATION: (213)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (214)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (228)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (231)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (267)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (292)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (300)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-1834

Query Match          10.4%; Score 149.2; DB 9; Length 309;
Best Local Similarity 95.6%; Pred. No. 3e-32;
Matches 195; Conservative 0; Mismatches 5; Indels 4; Gaps 4;

QY 322 ATCCGTCGATTCGACGGTTCCTCCTCAAGCCATGACCAAGATATCTCTGATTAATTC 381
DB 9 ATCCGTCGATTCGACGGTTCCTCCTCAAGCCATGACCAAGATATCTCTGATTAATTC 68
QY 382 CTAAGGAGTGGCTGCTGCCGCCCTGCTGCCCTGCTGCCCTGCTGCCCTGCTGCC 441
DB 59 CTAAGGAGTGGCTGCTGCCGCCCTGCTGCCCTGCTGCCCTGCTGCCCTGCTGCC 128
QY 442 TTG-ACAGGGTAAAGATTCTGTG-CTGAGGAGTGTGTGATGCG-ATTCTGGATAGATC 498
DB 129 TTGANAAGGGTAAAGATTCTGTG-CTGAGGAGTGTGTGATGCG-ATTCTGGATAGATC 521
QY 499 CCCACC-TTCCAAAGGGCTGAAAT 521
DB 189 CCCACC-TTCCAAAGGGCTGAAAT 212

RESULT 10
US-09-918-995-8879
; Sequence 8879, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8879
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(472)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-8879

Query Match          9.0%; Score 128.2; DB 10; Length 472;
Best Local Similarity 97.7%; Pred. No. 4.8e-26;
Matches 130; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1297 GAGTACTGCGATTAGGCGAGGCACTTTAATACCAACTGTAACTGTCTCAACTGTATAC 1356

US-09-930-440b-1.rnpb

Db 61 GAATTCCTGCAATTAGGCGAGCACTTTAATACCAAACTGTAACTGTCTCAACTGTATAC 120
QY 1357 AACTCAAAATACACACAGCTCATTTGGCTGCTCAGTCTTAACCTAGATGATGCTTTTGA 1416
DB 121 AACTCAAAATACACACAGCTCATTTGGCTGCTCAGTCTTAACCTAGATGATGCTTTTGA 180
QY 1417 ATTCATTTGCGATG 1429
DB 181 ATTCATTTGCGATG 193

RESULT 11
US-10-085-117-223
; Sequence 223, Application US/10085117
; Publication No. US2003023334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 223
; LENGTH: 177249
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(177249)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-223

Query Match          3.2%; Score 46.4; DB 16; Length 177249;
Best Local Similarity 45.6%; Pred. No. 1.3;
Matches 164; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

QY 952 ATTTCAGTGAATTCGAATAAACTCTCCTAGCAATGAAATCTCACAATTAAGCATTCAGG 1011
DB 71301 ACTTTACCTGCGAGTTTACTGGCTACTGTTCAAGAGAGAGTTCTTTGCTTATGACTGGCT 71360
QY 1012 TACCTTTTGTGAGCCTTAAAAAGTCTTATTTTGTAAGGGGCAAAAACCTTAGGAGTCAAC 1071
DB 71361 TACCTTAGATAATTCCTTAAAACTATATATATATATATATATATATACACACATACATAC 71420
QY 1072 AACTCTCAGTCAATTCATTTTCACAGATTTTTTTTGTGGAGAAATTCGTGTTTATATGATGA 1131
DB 71421 ACACACATATACATACACACACATATGATGATATATATATATATATATATATATATATG 71480
QY 1132 AATGGAATCAGAGGAAATTCGTAATGATTAATCCATCTGCTCTTTAGGAGCTCTCATT 1191
DB 71481 AATACACACATATATATATATATATATATATATATATATATATATATATATATATATAT 71540
QY 1192 ATCTCGGCTCTGCTGTTCCATTCCTATTTTAAAGTTGTCTAATTTTAAACCCTATAATA 1251
DB 71541 ATGTTATTAACAATATGATTTTAAATATAAAATCTTAATATATATATATATATATATATA 71600
QY 1252 TGTCTTCATTTTAAATATATATATTTTGGAAATCTAGGAAATCTGAGCTACTGCATTTA 1311
DB 71601 CATATACACATATATATATATATATATATATATATATATATATATATATATATATATATA 71660

RESULT 12
US-10-311-455-998
; Sequence 998, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENEROCK, Christian

```

```
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 998
; LENGTH: 6117
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: 4521
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-998

Query Match      3.2%; Score 45.4; DB 15; Length 6117;
Best Local Similarity 49.3%; Pred. No. 0.24; 146; Indels 1; Gaps 1;
Matches 144; Conservative 0; Mismatches 0;

QY 992 TCTCAGCAATAGCATTGAGTACCTTTTGTGAGCCTTAAAGCTTATTTTGTGAAGGG 1051
DB 316 TATTAGTATATGATTTATTTAGTGGTGGGAGGAATATAATTTGATTTTAGGTA 375
QY 1052 GCAAAATCTAGGAGTCACAACTCTCAGTCATTCATTTCACAGATTTTGTGGAGAA 1111
DB 376 GAAATATCTTTTCTATTATTTGTTTATTTTATTTTGTGAATTTATCGTGAAGTT 435
QY 1112 ATTTCTGTTTATGATGATGAATCGAATCAGAGGAAATTTGATTAATTCCTATC 1171
DB 436 AATATATTATTTTGGAGAGAGAGATTTAGAAATTTATGTTGTTATTAATTTAGTTA 495
QY 1172 TGCTTTAGGAGCTCTCATTAATCTCGGTCTCTGTTCTATCTATCTATTTAAAGTTGCT 1231
DB 496 GAAGTGTTGTTGGTTATT-GTAGGTATTTATTAATTAATTCGATTTTATGTTT 554
QY 1232 AATTTTAACCACTAATATATGCTTCACTTTTAATAATATTCATTTGGAA 1282
DB 555 TTATAGCTATATTTTATGTTGTTTAAATTTAATATGATGATTTGTA 605

RESULT 13
US-10-311-455-1924
; Sequence 1924, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1924
; LENGTH: 5771
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (214, 2796..2797, 4347)
US-10-221-613-33
Query Match      3.1%; Score 44.4; DB 13; Length 6109;
Best Local Similarity 52.1%; Pred. No. 0.47;
Matches 99; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1089 TTACAGATTTTGTGGAGAAATTTCTGTTTATATGATGAATGAATCAAGAGGAA 1148
DB 1641 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1700
QY 1149 AATTGTAATGATTAATTCATCTCTTTAGGAGCTCTCATTAATCTCGGTCTCGGTC 1208
DB 1701 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1760
QY 1209 CTAATCCTATTTAAGTTGCTGAATTTTAAACCACTAATATGCTCTCAATTTAATAA 1268
DB 1761 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1820
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1924

Query Match      3.1%; Score 44.8; DB 15; Length 5771;
Best Local Similarity 55.0%; Pred. No. 0.35;
Matches 88; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1079 AGTCATTTCATTCACAGATTTTGTGGAGAAATTTCTGTTTATTTGGAATCAATGGA 1138
DB 1803 AGTTTGTGTTTGTAGATATTTTGTGAACTAGTTTATATAATTAATTAAGGTTA 1852
QY 1139 TCAGAGGAAATTTGATTAATTCATTCATCTGTTTATGAGGAGCTCTCATTAATCTCGG 1198
DB 1863 AAAAGTAGATATTAGTAATTTTAAATGATGATATATGTTTATTATTATTATT 1922
QY 1199 TCTGCTTCTCTAATCTCTATTTTAAAGTTGTCTAAATTTA 1238
DB 1923 TGCGCTTTTATTATTATTATTATTATTATTAAAGTGA 1962

RESULT 14
US-10-221-613-33
; Sequence 33, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019258.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 33
; LENGTH: 6109
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (214, 2796..2797, 4347)
US-10-221-613-33

Query Match      3.1%; Score 44.4; DB 13; Length 6109;
Best Local Similarity 52.1%; Pred. No. 0.47;
Matches 99; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1089 TTACAGATTTTGTGGAGAAATTTCTGTTTATATGATGAATGAATCAAGAGGAA 1148
DB 1641 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1700
QY 1149 AATTGTAATGATTAATTCATCTCTTTAGGAGCTCTCATTAATCTCGGTCTCGGTC 1208
DB 1701 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1760
QY 1209 CTAATCCTATTTAAGTTGCTGAATTTTAAACCACTAATATGCTCTCAATTTAATAA 1268
DB 1761 TTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 1820
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:49 ; Search time 32.4251 seconds
(without alignments)
2004.184 Million cell updates/asc

Title: US-09-930-440B-2
Perfect score: 1159
Sequence: 1 MAFPKKQLQGLVAATIPMT.....GATGAGSVFVRDLSTLSLN 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 9.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A: Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1159	100.0	230	3	AAY96099 Human ald
2	1159	100.0	230	3	AAY93353 Human gly
3	1159	100.0	230	4	AAB84681 Amino aci
4	1159	100.0	230	6	AAO26544 Human ald
5	1104.5	95.3	284	5	ABB04624 Human di
6	1104.5	95.3	299	4	AAO13069 Human pol
7	1104.5	95.3	320	2	AAAY40417 A human N
8	1104.5	95.3	320	3	AAY79037 Human N-a
9	718.5	62.0	301	5	ABB57988 Polyadeny
10	675	58.2	147	4	AAAG81340 Human AFP
11	599	51.7	141	4	AAAG90590 Human imm
12	274.5	23.7	297	2	AAAR74749 Escherich
13	274.5	23.7	297	7	ADD5C650 Escherich
14	272.5	23.5	297	4	AAAB85682 Salmonell
15	254.5	22.0	237	4	AAAB96342 Putative
16	252.5	21.8	305	6	ABU01768 S. pneumo
17	252.5	21.8	305	6	ABP81408 Streptoco
18	247	21.3	305	5	ABP25875 Streptoco
19	242.5	20.9	233	6	ABM71755 Staphyloc
20	241.5	20.8	341	6	ADA33805 Acinetoba
21	239.5	20.7	302	7	ADC35653 E. faeciu
22	237.5	20.5	291	6	ABP79555 N. gonorr
23	233.5	19.3	289	6	ADB10102 Alloococ
24	222.5	19.2	301	2	AAAR40185 B. flavum
25	222.5	19.2	301	2	AAAR78235 Dihydrodi

26	222.5	19.2	301	2	AAW23283 Brevibact
27	222.5	19.2	301	2	AAW06582 B. lactof
28	222.5	19.2	301	2	AAW69549 Brevibact
29	222.5	19.2	301	2	AAW60501 Coryneb
30	222.5	19.2	301	2	AAW47403 B. lactof
31	222.5	19.2	301	2	AAW87740 Coryneb
32	222.5	19.2	301	4	AAW79661 Coryneb
33	222.5	19.2	301	4	AAE04859 Coryneb
34	222.5	19.2	301	4	AAU71889 C. glutam
35	222.5	19.2	301	4	AAU71889 C. glutam
36	222.5	19.2	301	6	AAU71889 C. glutam
37	222.5	19.2	301	6	AAU71889 C. glutam
38	222.5	19.2	309	4	AAU42223 Propionib
39	222.5	19.2	309	6	ABM38742 Propionib
40	219	18.9	298	6	ABU17664 Protein e
41	218.5	18.9	319	5	ABU15195 Protein e
42	218.5	18.9	1006	4	ABG17695 Novel hum
43	218.5	18.9	1131	4	ABG24897 Novel hum
44	218	18.8	309	6	ABU28508 Protein e
45	215	18.6	304	5	ABP25876 Streptoco

ALIGNMENTS

RESULT 1
AAY96099
ID AAY96099 standard; protein; 230 AA.
AC AAY96099;
XX
XX
DT 19-DEC-2000 (first entry)
XX
DE Human aldolase.
XX
KW Aldolase; human; sialylation; glycoprotein; plasminogen; transferrin;
KW thyrotropin; Na+,K+-ATPase.
XX
OS Hemo sapiens.
XX
PN WO200052135-A2.
XX
PD 08-SEP-2000.
XX
PF 01-MAR-2000; 2000WO-US005313.
XX
PR 02-MAR-1999; 99US-0122582P.
PR 08-DEC-1999; 99US-0169624P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
PA (UYWY-) UNIV WYOMING.
XX
PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
XX
XX MPI; 2000-572178/53.
XX
XX N-PSDB; AAA50567.

Recombinant production of sialylated glycoproteins using cells in which the expression of enzymes, e.g. sialic acid synthetase, involved in the sialylation reaction has been altered.

Claim 18; Page 100; 144pp; English.

The present sequence is that of human aldolase. The invention provides methods and recombinantly engineered cells for producing glycoproteins having sialylated oligosaccharides. The methods involve altering the expression of enzymes involved in carbohydrate processing. A claimed cell producing sialylated glycoprotein at above endogenous levels expresses at least 1 (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme that catalyzes the conversion of UDP-GlcNAc to ManNAc, sialic acid synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid transporter at above endogenous levels. Endogenous N-

CC acetylglucosaminidase activity may be suppressed. A claimed method for
 CC manipulating glycoprotein in an insect cell comprises enhancing the
 CC expression of 1 of the above enzymes, and a claimed method for producing
 CC sialylated glycoproteins involves expressing a heterologous protein
 CC (especially plasminogen, transferrin, Na⁺/K⁺-ATPase or thyrotropin) in
 CC the insect cell. Yeast, insect, fungal, plant and bacterial host cells
 CC can be engineered to produce new forms of sialylated glycoproteins,
 CC higher concentrations of sialylated glycoproteins and/or elevated
 CC concentrations of donor substrates (e.g. nucleotide sugars) required for
 CC sialylation
 XX
 SC Sequence 230 AA;

Query Match 100.0%; Score 1159; DB 3; Length 230;
 Best Local Similarity 100.0%; Pred. No. 8.1e-120;
 Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
 Db - MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
 QY 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAGIADGAVIAPFLKPEWT 120
 Db 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAGIADGAVIAPFLKPEWT 120
 QY 121 KDILINFLKEVAAAAPALPFPYVHIHVGALSLKESQELAQHAAGIADGAVIAPFLKPEWT 180
 Db 121 KDILINFLKEVAAAAPALPFPYVHIHVGALSLKESQELAQHAAGIADGAVIAPFLKPEWT 180
 QY 181 DFGQCVDQNRQQQAFALFGVDEQLLSALVWGATGAVGVSFVSRDLSTLLSN 230
 Db 181 DFGQCVDQNRQQQAFALFGVDEQLLSALVWGATGAVGVSFVSRDLSTLLSN 230

RESULT 2
 AAY93353
 ID AAY90353 standard; protein; 230 AA.
 XX AAY90353;
 AC
 DT 04-DEC-2000 (first entry)
 DE Human glycosylation enzyme clone HDPAX85 protein sequence.
 XX
 KW Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
 KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
 KW haematopoietic cell deficiency; blood coagulation disorder; asthma;
 KW afibrinogenemia; blood platelet disorder; thrombocytopenia; neoplasia;
 KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
 KW allergic encephalomyelitis; allergic reaction; organ rejection;
 KW graft-versus-host disease; inflammation; hyperproliferative disorder;
 KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase.
 XX
 OS Homo sapiens.
 XX WO200052136-A2.
 XX
 PD 08-SEP-2000.
 XX
 PF 01-MAR-2000; 2000WO-US005325.
 XX
 PR 02-MAR-1999; 99US-0122409P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Coleman TA;
 XX
 DR WPI; 2000-572179/53.
 DE N-PSDB; AAA37764.
 XX
 PT New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
 PT synthetase, sialic acid synthetase and aldolase and nucleic acids
 PT encoding the proteins for treating e.g., immune system disorders,

PT microbial diseases.
 XX
 PS Claim 12; Page 113-114; 115pp; English.
 XX
 CC This sequence represents a human glycosylation enzyme clone of the
 CC invention, designated HDPAX85. This protein clone is an aldolase. The
 CC sequences are useful as reagents for the differential identification of
 CC the tissues or cell types present in a biological sample, as
 CC immunological probes, for treating a disease or condition, and as a
 CC under expression of such polypeptide, for the detection and/or treatment
 CC of disorders involving aberrant glycolysis, e.g. cramps, myoglobinuria,
 CC and as tumour marker and/or immunotherapy targets. They may also be used
 CC to differentiate, proliferate and attract cells leading to the
 CC regeneration of tissues, to modulate mammalian characteristics (e.g. in
 CC cosmetic surgery) or mammalian metabolism affecting catabolism, anabolism
 CC processing, utilisation and energy storage, to change a mammal's mental
 CC state by influencing and as a food additive or preservative. The proteins
 CC can be used to assay protein levels in a sample, as a marker or detector
 CC of an immune system disorder, to inhibit cytokine activity, and as a
 CC vaccine. They may further be used to treat immune system or of
 CC haematopoietic cell deficiencies or disorders, blood coagulation
 CC disorders (e.g. afibrinogenemia), blood platelet disorders (e.g.
 CC thrombocytopenia), wounds resulting from trauma or surgery, autoimmune
 CC disorders (e.g. Addison's disease, multiple sclerosis, allergic
 CC encephalomyelitis), allergic reactions (e.g. asthma), organ rejection,
 CC graft-versus-host disease, inflammation, hyperproliferative disorders
 CC (e.g. neoplasia, purpura, sarcoidosis), diseases caused by viruses (e.g.
 CC hepatitis, meningitis, AIDS), bacteria and fungi (include e.g.
 CC tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia, cellulitis),
 CC and diseases caused by parasites (e.g. amoebiasis, coccidiosis,
 CC leishmaniasis, scabies, malaria, toxoplasmosis)
 XX
 SQ Sequence 230 AA;

Query Match 100.0%; Score 1159; DB 3; Length 230;
 Best Local Similarity 100.0%; Pred. No. 8.1e-120;
 Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
 Db 1 MAFPKKQLGLVAATITPTMTENGESVIGQYDYLKQGVKNIFVNGTIGGLSLSV 60
 QY 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAGIADGAVIAPFLKPEWT 120
 Db 61 SERQVAEWTGKOKLDQVLIHVGLSLKESQELAQHAAGIADGAVIAPFLKPEWT 120
 QY 121 KDILINFLKEVAAAAPALPFPYVHIHVGALSLKESQELAQHAAGIADGAVIAPFLKPEWT 180
 Db 121 KDILINFLKEVAAAAPALPFPYVHIHVGALSLKESQELAQHAAGIADGAVIAPFLKPEWT 180
 QY 181 DFGQCVDQNRQQQAFALFGVDEQLLSALVWGATGAVGVSFVSRDLSTLLSN 230
 Db 181 DFGQCVDQNRQQQAFALFGVDEQLLSALVWGATGAVGVSFVSRDLSTLLSN 230

RESULT 3
 AAB84681
 ID AAB84681 standard; protein; 230 AA.
 XX AAB84681;
 AC
 DT 17-SEP-2001 (first entry)
 DE Amino acid sequence of a human aldolase.
 XX
 KW Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
 KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
 KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
 KW vaccine.
 XX
 OS Homo sapiens.
 XX WO200142492-A1.
 PN

XX PD 14-JUN-2001.
XX PF 07-DEC-2000; 2000WO-US033136.
XX PF 09-DEC-1999; 99US-0169839P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (UJJO) UNIV JOHNS HOPKINS.
PA (UTEM) UNIV TEMPLE.
PA (UWY-) UNIV WYOMING.
XX
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;
PI Jarvis D;
XX
XX WPI; 2001-441575/47.
DR N-PSDB; AAM28456.
XX
XX Cells producing cytidine monophosphate-sialic acid and sialylated
PT glycoprotein above endogenous levels for production of vaccines and
PT therapeutics.
XX
XX Claim 18; Fig 28; 182pp; English.
XX
XX The specification describes a method for manipulating carbohydrate
CC processing pathways in cells of interest. The methods are used to
CC manipulate multiple pathways involved with the sialylation reaction by
CC using recombinant DNA technology and substrate feeding approaches to
CC enable the production of sialylated glycoproteins in the cells. The
CC sialylation process involves the post-translational addition of the donor
CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific
CC acceptor carbohydrate. The cells express at least one enzyme, selected
CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,
CC CMP-SA synthetase and CMP-SA transporter, where the expression of each enzyme is
CC enhanced to above endogenous levels. The novel method is useful for
CC manipulating glycoprotein production in an insect cell. Further methods
CC of the invention are useful for producing sialylated glycoproteins. The
CC sialylated glycoprotein produced by the above mentioned methods are
CC useful as pharmaceutical compositions, vaccines, diagnostics and
CC therapeutics. This sequence represents the human aldolase protein of the
XX invention
XX
XX Sequence 230 AA;
XX
XX Query Match 100.0%; Score 1159; DB 4; Length 230;
XX Best Local Similarity 100.0%; Pred. No. 8.1e-120;
XX Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MAFPKKLOGLVAATITPTMTENGINFVIGQYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
DB 1 MAFPKKLOGLVAATITPTMTENGINFVIGQYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
OY 61 SERQVAEEWTKGDKLDQVLIHVGLSLKESQELAHAAEIGADGIAVIAPIFLKPT 120
DB 61 SERQVAEEWTKGDKLDQVLIHVGLSLKESQELAHAAEIGADGIAVIAPIFLKPT 120
OY 121 KDILINFLKEVAAPALPFYTHIPALTYKIRAEELLDGILDKIPTFQGLKESDITLL 180
DB 121 KDILINFLKEVAAPALPFYTHIPALTYKIRAEELLDGILDKIPTFQGLKESDITLL 180
OY 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230
DB 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230
RESULT 4
AAO26544
ID AAO26544
XX AAO26544 standard; protein; 230 AA.
AC AAO26544;

XX DT 06-MAR-2003 (first entry)
XX DE Human aldolase protein.
XX KW Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAC; sialic acid;
KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
KW transporter; sialylated glycoprotein; human; aldolase.
XX OS Homo sapiens.
XX
XX US2002142386-A1.
XX PD 03-OCT-2002.
XX
XX 16-AUG-2001; 2001US-00930440.
XX
XX 02-MAR-1999; 99US-0122582P.
XX 08-DEC-1999; 99US-0169839P.
XX 25-AUG-2000; 2000US-0227579P.
XX
XX (BETE/) BETENBAUGH M J.
XX (LAWR/) LAWRENCE S.
XX (LEEY/) LEE Y C.
XX (COLE/) COLEMAN T A.
XX
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
XX WPI; 2003-182519/09.
XX N-PSDB; AAL53991.
XX
XX Manipulating glycoprotein production in insect cell, involves enhancing
PT expression of enzymes involved in carbohydrate processing pathway such as
PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
XX
XX Claim 18; Fig 28; 88pp; English.
XX
XX The invention relates to a novel method for manipulating glycoprotein
CC production in an insect cell comprising enhancing expression of an
CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
CC catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, sialic acid
CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
CC synthetase or CMP-SA transporter, where the expression of each enzyme is
CC enhanced to above endogenous levels. The novel method is useful for
CC manipulating glycoprotein production in an insect cell. Further methods
CC of the invention are useful for producing sialylated glycoproteins. The
CC sialylated glycoprotein produced by the above mentioned methods are
CC useful as pharmaceutical compositions, vaccines, diagnostics and
CC therapeutics. This sequence represents the human aldolase protein of the
XX invention
XX
XX Sequence 230 AA;
XX
XX Query Match 100.0%; Score 1159; DB 6; Length 230;
XX Best Local Similarity 100.0%; Pred. No. 8.1e-120;
XX Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MAFPKKLOGLVAATITPTMTENGINFVIGQYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
DB 1 MAFPKKLOGLVAATITPTMTENGINFVIGQYVDYLKKEQGVKNIFVNGTTGEGLSLV 60
OY 61 SERQVAEEWTKGDKLDQVLIHVGLSLKESQELAHAAEIGADGIAVIAPIFLKPT 120
DB 61 SERQVAEEWTKGDKLDQVLIHVGLSLKESQELAHAAEIGADGIAVIAPIFLKPT 120
OY 121 KDILINFLKEVAAPALPFYTHIPALTYKIRAEELLDGILDKIPTFQGLKESDITLL 180
DB 121 KDILINFLKEVAAPALPFYTHIPALTYKIRAEELLDGILDKIPTFQGLKESDITLL 180
OY 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230
DB 181 DFGQCVQDNRQQQPAFLFGVDEQLLSALVMGATGAVGSFVSRDLSTLLN 230

RESULT 5
AB04624
ID A3304624 standard; protein; 284 AA.
AC ABB04624;
XX
XX 04-MAR-2002 (first entry)
XX
XX Human dihydrodipyrindine synthetase 31 protein SEQ ID NO:2.
DE
XX Human; dihydrodipyrindine synthetase 31; malignant tumour; nosohaemia;
KW HIV infection; immunological disease; inflammation.
XX
XX Homo sapiens.
OS
XX
XX CN1307105-A.
PN
XX
XX 08-AUG-2001.
PD
XX
XX 28-JAN-2000; 2000CN-00111575.
PF
XX
XX 28-JAN-2000; 2000CN-00111575.
PR
XX
XX (BODA-) BODAO GENE TECH CO LTD SHANGHAI.
PA
XX
XX Mao Y, Xie Y;
PI
XX
XX WPI; 2002-026890/04.
DR
XX
XX N-PSDB; ABA24275.
DR
XX
XX New polypeptide for treating malignant tumor and HIV infection, comprises
PT the polypeptide-dihydrodi pyridine synthetase 31 and polynucleotide for
PT coding said polypeptide.
PT
XX
XX Claim 1; Page 26-27 (Disclosure); 34pp; Chinese.
PS
XX
XX The present sequence represents human dihydrodipyrindine synthetase 31
CC (-). (I) can be used in the treatment of various diseases, such as
CC malignant tumour, nosohaemia, HIV infection, immunological diseases and
CC inflammations. The present invention also describes the antagonist
CC resisting (I) and its treatment effect
CC
XX Sequence 284 AA;
SQ
Query Match 95.3%; Score 1104.5; DB 5; Length 284;
Best Local Similarity 91.0%; Pred. No. 1.2e-113;
Matches 222; Conservative 5; Mismatches 5; Indels 15; Gaps 1;
QY 1 MAPPKKQLQGLVAATITPMTENGSEINFSVIGQVDYLVKEQGVKNIPVNGTTGEGLSLV 60
DB 1 MAPPKKQLQGLVAATITPMTENGSEINFSVIGQVDYLVKEQGVKNIPVNGTTGEGLSLV 60
QY 61 SERQVAEEWVTGKDKLDQVITHVGALSKESQELAQHAAEIGADGIAVIAPFELKPWT 120
DB 61 SERQVAEEWVTGKDKLDQVITHVGALSKESQELAQHAAEIGADGIAVIAPFELKPWT 120
QY 121 KDLINFLKEVAAAAPALPYYHYHIPALTGKIRAEELDGLDKIPTFQGLKFSSTDLL 180
DB 121 KDLINFLKEVAAAAPALPYYHYHIPALTGKIRAEELDGLDKIPTFQGLKFSSTDLL 180
QY 181 DFGCCVDQNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 225
DB 181 DFGCCVDQNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 225
QY 226 TLIS 229
DB 241 LALN 244
RESULT 6
AA013069

AA013069 standard; protein; 299 AA.
XX
XX AA013069;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 26961.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX
XX WO200164835-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 26-FEB-2001; 2001WO-US004927.
PF
XX
XX 28-FEB-2000; 2000US-00515126.
PR
XX
XX 18-MAY-2000; 2000US-00577409.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-514838/56.
DR
XX
XX N-PSDB; AA193000.
DR
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
PT
XX
XX Claim 20; SEQ ID NO 26961; 1399pp + Sequence Listing; English.
PS
XX
XX The invention relates to human polynucleotides (AA179941-AA193041) and
CC the encoded proteins (AA000010-AA03910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 299 AA;
SQ
Query Match 95.3%; Score 1104.5; DB 4; Length 299;
Best Local Similarity 91.0%; Pred. No. 1.3e-113;
Matches 222; Conservative 5; Mismatches 5; Indels 15; Gaps 1;
QY 1 MAPPKKQLQGLVAATITPMTENGSEINFSVIGQVDYLVKEQGVKNIPVNGTTGEGLSLV 60
DB 6 MAPPKKQLQGLVAATITPMTENGSEINFSVIGQVDYLVKEQGVKNIPVNGTTGEGLSLV 65
QY 61 SERQVAEEWVTGKDKLDQVITHVGALSKESQELAQHAAEIGADGIAVIAPFELKPWT 120
DB 61 SERQVAEEWVTGKDKLDQVITHVGALSKESQELAQHAAEIGADGIAVIAPFELKPWT 125
QY 121 KDLINFLKEVAAAAPALPYYHYHIPALTGKIRAEELDGLDKIPTFQGLKFSSTDLL 180
DB 126 KDLINFLKEVAAAAPALPYYHYHIPALTGKIRAEELDGLDKIPTFQGLKFSSTDLL 185
QY 181 DFGCCVDQNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 225
DB 186 DFGCCVDQNRQQQFAFLFGVDEQLLSALVMGATGAVGS-----FVSRDLS 245
QY 226 TLIS 229

Db	246 LALN 249	Matches 222; Conservative 2; Mismatches 5; Indels 15; Gaps 1;
RESULT 7		
AA40417		
ID	AA40417 standard; protein; 320 AA.	
XX		
AC	AA40417;	
XX		
DT	26-NOV-1999 (first entry)	
XX		
DE	A human N-acetylneuraminase lyase (hNANL) protein.	
XX		
KW	Human; N-acetylneuraminase lyase; NANL; N-acetylneuraminase aldolase;	
KW	acyneuraminase-pyruvate lyase; sialic acid; degradation;	
KW	energy metabolism; sialic acid metabolism; gangliosidosis; Tay-Sachs;	
KW	beta-hexosaminidase deficiency; beta-galactosidase deficiency; cancer;	
KW	autoimmune disorder; acquired immune deficiency syndrome; AIDS;	
KW	Addison's disease; Crohn's disease; diabetes mellitus; atherosclerosis;	
KW	asthma.	
XX		
OS	Homo sapiens.	
XX		
PN	US5962302-A.	
XX		
PD	05-OCT-1999.	
XX		
PF	20-FEB-1998; 98US-00027013.	
XX		
PR	20-FEB-1998; 98US-00027013.	
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Hillman JL, Corley NC, Shah P;	
XX		
DR	WPI; 1999-571273/48.	
DR	N-PSDB; AA222092.	
XX		
PT	Nucleic acids encoding human N-acetylneuraminase lyase useful for	
PT	treating GM1 and GM2 gangliosidosis (such as Tay-Sachs/beta-	
PT	hexosaminidase deficiency and beta-galactosidase deficiency).	
XX		
PS	Claim 1; Fig 1A-F; 31pp; English.	
XX		
CC	The present sequence represents a human N-acetylneuraminase lyase (hNANL)	
CC	protein. NANL (also called N-acetylneuraminase aldolase and	
CC	acyneuraminase-pyruvate lyase) is an enzyme that catalyses the	
CC	conversion of N-acetylneuraminic acid (sialic acid) in pyruvate and N-	
CC	acetyl-D-mannosamine (and the reverse reaction). In bacteria NANL is	
CC	involved in energy metabolism as the degradation of sialic acid produces	
CC	usable carbon and energy sources. The expression of sialic acid is	
CC	regulated by the level of sialic acid. NANL is found in a number of	
CC	mammalian tissues and is a key enzyme in the metabolism of sialic acid.	
CC	hNANL polypeptides and polynucleotides may be used in the diagnosis,	
CC	prevention and treatment of disorders associated with defective sialic	
CC	acid metabolism, e.g. gangliosidosis (such as Tay-Sachs/beta-	
CC	hexosaminidase deficiency and beta-galactosidase deficiency). They may	
CC	also be used to study the expression and function of NANL proteins and	
CC	their role in sialic acid metabolism. The hNANL proteins may then be used	
CC	as antigens in the production of antibodies and in assays to identify	
CC	modulators of NANL expression and activity. These antagonists of NANL may	
CC	be used to down regulate NANL expression and activity which may prove	
CC	useful in the treatment of disorders associated with over expression and	
CC	activity of NANL such as cancers (e.g. leukemia, melanoma and cancers of	
CC	the lung, brain, prostate, ovary and pancreas) and autoimmune disorders	
CC	(e.g. acquired immune deficiency syndrome (AIDS), Addison's disease,	
CC	Crohn's disease, diabetes mellitus, atherosclerosis and asthma)	
XX		
SQ	Sequence 320 AA;	
Query Match	95.3%; Score 1104.5; DB 2; Length 320;	
Best Local Similarity	91.0%; Pred. No. 1.5e-113;	
QY	1 MAPEKKKLGVAATITPMTENGSEINFVSIGQYVDYLWKQGVKNIFVNGTTGEGLSLV 60	
DB	1 MAPEKKKLGVAATITPMTENGSEINFVSIGQYVDYLWKQGVKNIFVNGTTGEGLSLV 60	
QY	61 SERQVAEEWTKGDKLDQVLIHVGALSKESELAQAHAIEIGADGIATAPFLKPT 120	
DB	61 SERQVAEEWTKGDKLDQVLIHVGALSKESELAQAHAIEIGADGIATAPFLKPT 120	
QY	121 KDILINFLKEVAAPALPFVYHIPALTGVKIPAEELLDGILDKIPFQSKESDTELL 180	
DB	121 KDILINFLKEVAAPALPFVYHIPALTGVKIPAEELLDGILDKIPFQSKESDTELL 180	
QY	181 DFGQVDQNFQQQFAFLGVDEQLLSALVMGATGAVGS-----FVSRLS 225	
DB	181 DFGQVDQNFQQQFAFLGVDEQLLSALVMGATGAVGS-----FVSRLS 225	
QY	226 TLIS 229	
DB	241 LALN 244	
RESULT 8		
AA479037		
ID	AA479037 standard; protein; 320 AA.	
XX		
AC	AA479037;	
XX		
DT	20-JUN-2000 (first entry)	
XX		
DE	Human N-acetylneuraminase lyase (hNANL) amino acid sequence.	
XX		
KW	Cancer; autoimmune disease; diagnosis; treatment; leukaemia; lymphoma;	
KW	melanoma; myeloma; teratocarcinoma; sarcoma; AIDS; Addison's disease;	
KW	allergy; asthma; atherosclerosis; bronchitis; diabetes mellitus; anaemia;	
KW	Graves' disease; irritable bowel syndrome; Myasthenia gravis;	
KW	rheumatoid arthritis; osteoarthritis; Sjogren's syndrome; infection;	
KW	human N-acetylneuraminase lyase; hNANL.	
XX		
OS	Homo sapiens.	
XX		
PN	US6030824-A.	
XX		
PD	29-FEB-2000.	
XX		
PF	01-FEB-1999; 99US-00244233.	
XX		
PR	20-FEB-1998; 98US-00027013.	
XX		
PA	(INCY-) INCYTE PHARM.	
XX		
PI	Shah P, Corley NC, Hillman JL;	
XX		
DR	WPI; 2000-223164/19.	
DR	N-PSDB; AA398699.	
XX		
PT	New human N-acetylneuraminase lyase useful in the diagnosis, treatment	
PT	and prevention of cancer and autoimmune disorders, comprises of 320 amino	
PT	acid residues.	
XX		
PS	Claim 1; Fig 1; 31pp; English.	
XX		
CC	This sequence represents a human N-acetylneuraminase lyase (hNANL) amino	
CC	acid sequence. hNANL catalyses the conversion of N-acetylneuraminic acid	
CC	(sialic acid) into pyruvate and N-acetyl-D-mannosamine, as well as the	
CC	reverse reaction. Sialic acid is an essential constituent of gangliosides	
CC	a class of glycolipids. Gangliosides are components of the cell membrane	
CC	and act as receptors and are involved in cell cell interactions, and	
CC	signal transduction. The hNANL protein sequence, and the nucleotide	
CC	sequence encoding it are useful in the diagnosis, treatment and	
CC	prevention of cancer and autoimmune disease. The cancers diagnosed,	
CC	treated and prevented include adenocarcinoma, leukaemia, lymphoma,	

CC melanoma, myeloma, teratocarcinoma, sarcoma and in particular cancers of
CC the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix,
CC gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver,
CC lung, muscle, ovary, pancreas, penis, prostate, salivary glands, spleen,
CC skin, uterus, thymus and thyroid. The autoimmune disorders diagnosed,
CC treated and prevented include AIDS, Addison's disease, adult respiratory
CC distress syndrome, allergies, asthma, atherosclerosis, bronchitis,
CC diabetes mellitus, anaemia, Graves' disease, irritable bowel syndrome,
CC myasthenia gravis, rheumatoid arthritis, osteoarthritis, Sjogren's
CC syndrome, viral, bacterial, fungal, parasitic, protozoal and helminthic
CC infections

XX SQ Sequence 320 AA;

Query Match 95.3%; Score 1104.5; DB 3; Length 320;
Best Local Similarity 91.0%; Pred. No. 1.5e-113; Indels 15; Gaps 1;
Matches 222; Conservative 2; Mismatches 5;

QY 1 MAFPKKLGVLVAATITPTMTENGESVIGQYVDYLVKEQVKNIFVNGTTGEGLSLV 60
DB 1 MAFPKKLGVLVAATITPTMTENGESVIGQYVDYLVKEQVKNIFVNGTTGEGLSLV 60

QY 61 SERQVABEWTKGKDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPT 120
DB 61 SERQVABEWTKGKDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPT 120

QY 121 KDILINFLKEVAAPALPFYVYHPALTGVKIRAEELDGLDKIPTFGKFSDDTL 180
DB 121 KDILINFLKEVAAPALPFYVYHPALTGVKIRAEELDGLDKIPTFGKFSDDTL 180

QY 181 DFGQVQNRQQAFAFLGVDEQLLSALVMGATGAVGS-----FVSRDL 225
DB 181 DFGQVQNRQQAFAFLGVDEQLLSALVMGATGAVGS-----FVSRDL 225

QY 226 TLIS 229
DB 241 LALN 244

RESULT 9
ABB97988
ID ABB97988 standard; protein; 301 AA.
XX AC ABB97988;
XX DT 02-OCT-2002 (first entry)
XX DE Polyadenylate binding protein 33.11.
XX KW Polyadenylate binding 33.11; embryo development deformity; tumour;
XX KW protein metabolism disorder.
XX OS Unidentified.
XX PN CN1340520-A.
XX PD 20-MAR-2002.
XX PF 31-AUG-2000; 2000CN-00119821.
XX PR 31-AUG-2000; 2000CN-00119821.
XX PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX PI Mao Y, Xie Y;
XX DR WPI; 2002-436414/47.
XX DR N-PSDB; ABN86096.

XX PT Polypeptide-polyadenylate binding 33.11 and polynucleotide for coding it.
XX PS Claim 1; Page 29 (disclosure); 35pp; Chinese.
XX

CC The invention relates to a novel polyadenylate binding protein 33.11, the
CC polynucleotide encoding it, and the process for preparing the polypeptide
CC by DNA recombination techniques. The polypeptide is used in treating
CC diseases such as embryo development deformity, tumour and protein
CC metabolism disorder. The current sequence represents the polyadenylate
CC binding protein 33.11 of the invention

XX SQ Sequence 301 AA;

Query Match 62.0%; Score 718.5; DB 5; Length 301;
Best Local Similarity 85.8%; Pred. No. 9.4e-71;
Matches 145; Conservative 4; Mismatches 5; Indels 15; Gaps 1;

QY 76 DKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPTKDILINFLKEVAAA 135
DB 76 ERLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFLKPTKDILINFLKEVAAA 116

QY 136 PALPFYVYHPALTGVKIRAEELDGLDKIPTFGKFSDDTLDFGQVQNRQQA 195
DB 117 PALPFYVYHPALTGVKIRAEELDGLDKIPTFGKFSDDTLDFGQVQNRQQA 176

QY 196 FLFGVDEQLLSALVMGATGAVGS-----FVSRDLSTLIS 229
DB 177 FLFGVDEQLLSALVMGATGAVGS-----FVSRDLSTLIS 225

RESULT 10
AAG81340
ID AAG81340 standard; protein; 147 AA.
XX AC AAG81340;
XX DT 10-SEP-2001 (first entry)
XX DE Human AFP protein sequence SEQ ID NO:198.
XX KW Human; secreted protein; secretion; bacterial cell; fungal cell;
XX KW eukaryotic cell; fusion protein; maltose binding protein;
XX KW immunoglobulin constant region; polyhistidine tag.
XX OS Homo sapiens.
XX PN WO200129221-A2.
XX PD 26-APR-2001.
XX PF 20-OCT-2000; 2000WO-US029052.
XX PR 20-OCT-1999; 99US-0150712P.
XX PA (ZYMO) ZYMOGENETICS INC.
XX PI Conklin DC, Yee DP;
XX DR WPI; 2001-300340/31.
XX DR N-PSDB; AAH52191.

XX PT Isolated polypeptide for directing secretion of proteins of interest from
XX PT a host cell including, e.g. bacteria, includes contiguous amino acid
XX PT residues of polypeptide with specified amino acids.
XX PS Claim 1; Page 348-349; 617pp; English.
XX CC AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242
XX CC to AAG81453. The secreted proteins can be used for directing the
XX CC secretion of proteins of interest from a host cell including bacteria,
XX CC fungal cells, and cultured higher eukaryotic cells. The present invention
XX CC also describes fusion proteins, where a secreted protein of the invention
XX CC is operably linked via a peptide bond or peptide linker to a second
XX CC protein selected from the group consisting of maltose binding protein, an
XX CC immunoglobulin constant region, a polyhistidine tag and a peptide given
XX CC in AAG81453

```
SQ Sequence 147 AA;
Query Match 58.2%; Score 675; DB 4; Length 147;
Best Local Similarity 93.2%; Pred. No. 2.2e-66;
Matches 138; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 MAFPKKLGKGLVAATITPMTENGESVIGQYVDYLKQGVKNIFVNGTGTGEGLSLSV 60
   |||||
Db 1 MAFPKKLGKGLVAATITPMTENGESVIGQYVDYLKQGVKNIFVNGTGTGEGLSLSV 60
   |||||

QY 61 SERQVAEWTGKDKLDQVLIHWGALSLSKESQELAQHAAEIGADGIAVIAPFFLKPT 120
   |||||
Db 61 SERQVAEWTGKDKLDQVLIHWGALSLSKESQELAQHAAEIGADGIAVIAPFFLKPT 120
   |||||

QY 121 KDILINFLKVAAPALPFFYYHIPAL 148
   |||||
Db 121 KDILINFLKVAAPALPFFYYHIPAL 148
   |||||

RESULT 11
AAM90590 ID AAM90590 standard; protein; 141 AA.
XX AC AAM90590;
XX DT 07-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen SEQ ID NO:18183.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KY cytotatic; gene therapy; vaccine; metastasis.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001354.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186309P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190078P.
PR 18-APR-2000; 2000US-0195123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227609P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
```

```

PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249248P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259675P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX N-PSDB; AAK63371.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Claim 11; SEQ ID NO 18183; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK62170 to AAK61921. (I) have cytotoxic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK62169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 141 AA;
XX
XX Query Match 51.7%; Score 599; DB 4; Length 141;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-58;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 102 EIGADGIAVAPFLKFWTKDIINFLKEVAAPALPFYVYHIALTGKIRAEELLOG 161
XX
XX 3 EIGADGIAVAPFLKFWTKDIINFLKEVAAPALPFYVYHIALTGKIRAEELLOG 62
XX
XX 162 ILDKIPTFQGLKFSOTLLDFFGQVQDNROQQAFLFGVDEQLLSALVMGATGAVGS 218
XX

```

```

Db 63 ILDKIPTFQGLKFSOTLLDFFGQVQDNROQQAFLFGVDEQLLSALVMGATGAVGS 119
RESULT 12
AAR74749
ID AAR74749 standard; protein; 297 AA.
XX
XX AAR74749;
XX
XX 10-JAN-1996 (first entry)
XX
XX Escherichia coli N-acetylneuraminic acid aldolase.
XX
XX Escherichia coli; N-acetylneuraminic acid aldolase; sialic acid;
XX recombinant production; carcinoma diagnosis.
XX
XX Escherichia coli.
XX JP07107968-A.
XX
XX 25-APR-1995.
XX
XX 14-OCT-1993; 93JP-00256903.
XX
XX 14-OCT-1993; 93JP-00256903.
XX
XX (TOYM) TOYOB0 KK.
XX
XX WPI; 1995-130169/25.
XX
XX N-PSDB; AAK92410.
XX
XX Recombinant proctn. of N-acetylneuraminic acid aldolase - in gram-negative
XX bacterium other than Escherichia sp., useful for measuring change in
XX sialic acid content for diagnosis of carcinoma.
XX
XX Claim 1; Page 7-8; 13pp; Japanese.
XX
XX AAK92410 encodes AAR74749 Escherichia coli N-acetylneuraminic acid
XX aldolase, which can be recombinantly produced in other Gram negative
XX bacteria. The enzyme is useful for measuring changes in sialic acid
XX content, and is therefore useful in the diagnosis of carcinomas
XX
XX Sequence 297 AA;
XX
XX Query Match 23.7%; Score 274.5; DB 2; Length 297;
XX Best Local Similarity 34.3%; Pred. No. 1.8e-21;
XX Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;
XX
XX 8 LQGLVAATITPMTENGSEINFSVIGQVYDLVKKGQVKNIQVNGTGTGEGLSLSVSERRQ-- 65
XX
XX 5 LRGVMAALITFPQQQALDKASRLRLVQFWI--QQGIDGLYVGSGTGEAFVCSLSEREQVL 63
XX
XX 66 --VAREWTKGDKLDQVIVHVGALSLSKESQELAQHAAEIGADGIAVAPFLKFWTKDI 123
XX
XX 64 EIVAREE--AKGKIKL--IAHVGCSTAESQQAASAKRYGFDVAVSATVPFY--PFSFEE 117
XX
XX 124 LINFLEKVAAPALPFYVYHIALTGKIRAEELGILDKIPTFQGLKFSOTLLDFFG 183
XX
XX 118 HCDHYRAIDISAGLPMVYVNIIFALSGVKLTLDQI--NTLVLPFGVGLAKQTSGLDQME 175
XX
XX 184 QCVQDNROQQAFLF-GVDEQLLSALVMGATGAVGS 218
XX
XX 176 QI---RREHPDVLVLYNGYDEIFASGLLAGADGGIGS 208
XX
RESULT 13
ADD90650
ID ADD90650 standard; protein; 297 AA.
XX
XX ADD90650;
XX
XX 29-JAN-2004 (first entry)
XX

```

DE Escherichia coli protein. SEQ ID NO:3.
 XX microorganism; N-acetylneuraminic acid; Escherichia coli.
 XX Escherichia coli.

XX WC2003072783-A1.

XX 04-SEP-2003.

XX 28-FEB-2003; 2003WO-JP002332.

XX 28-FEB-2002; 2002JP-00053696.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Ishikawa M, Koizumi S;

XX WPI; 2003-721776/68.

XX N-PSDB; ADD90651.

XX Microbial production of N-acetylneuraminic acid for use e.g. in chemical

XX synthesis or as pharmaceutical intermediates.

XX Disclosure; SEQ ID NO 3; 52pp; Japanese.

XX The present invention describes a process for producing N-
 CC acetylneuraminic acid by culturing a microorganism capable of producing N-
 CC acetylneuraminic acid but with no or lower activity in decomposing N-
 CC acetylneuraminic acid as compared with the wild-type strain, in a culture
 CC medium to accumulate the product for isolation. The method can be used
 CC for the production of N-acetylneuraminic acid for use e.g. in chemical
 CC synthesis or as pharmaceutical intermediates. The process is efficient
 CC and economical, in which a microorganism used has no or lower activity in
 CC decomposing N-acetylneuraminic acid as compared with the wild-type
 CC strain, from cheap starting materials and with simple operation. The
 CC present sequence represents an Escherichia coli protein, which is used in
 CC the exemplification of the present invention.

XX Sequence 297 AA;

XX Query Match 23.7%; Score 274.5; DB 7; Length 297;
 XX Best Local Similarity 34.3%; Pred. No. 1.8e-21;
 XX Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;
 QY 8 LQGLVAATITMTENGSENFVIGQYVDLVKQGVKAFVNGTGTGEGLSLSVSRQ-- 65
 DB 5 LRGVMAALITPDQQALDKASLRLLVDFNI--QGGIDGLVGGSTGEAFVQSLSRQVL 63
 QY 66 --VAPFWTKGDKLDQVIVHVGALSLSKESQELAQHAAEIGADGIAVIAPFLKPTWKDI 123
 DB 64 EIVAE--AKGKIKL--LAHVGCYSTAESQQAASAKRYGFDVAVSATVPFFY-PFSFE 117
 QY 124 LINFKEVAAAPALPFFYYHPIALTGVKIRAEELLDGILDKIPTFGKLFSDTDLRFG 183
 DB 118 HCDHYRAIDSADGLEWVYVNIIPALSGVKLTLDQI--NTLVLPGVGALKQTSGLDYQWE 175
 QY 184 QCVQDNQQQFAFLF-GVDEQLLSALVWGATGAVGS 218
 DB 176 QI---RRHPDLVLYNGYDEIFASGLLAGADGGIGS 208

RESULT 14

AAB85682

ID AAB85682 standard; protein; 297 AA.

XX AAB85682;

XX 29-OCT-2001 (first entry)

XX Salmonella virulent protein.

XX Salmonella; virulent; medicament; infection; Gram-negative bacteria;

KW gastro-enteritis; veterinary; antimicrobial; antibacterial; vaccine;
 XX antiinflammatory.
 OS Salmonella typhimurium.

PN WC200157075-A2.

XX 09-AUG-2001.

XX 02-FEB-2001; 2001WO-GB000449.

XX 03-FEB-2000; 2000GB-00002552.

XX 03-FEB-2000; 2000GB-00002554.

XX 03-FEB-2000; 2000GB-00002555.

XX 03-FEB-2000; 2000GB-00002556.

XX 03-FEB-2000; 2000GB-00002557.

XX 03-FEB-2000; 2000GB-00002558.

XX 03-FEB-2000; 2000GB-00002559.

XX 03-FEB-2000; 2000GB-00002560.

XX 03-FEB-2000; 2000GB-00002561.

XX 03-FEB-2000; 2000GB-00002562.

XX 03-FEB-2000; 2000GB-00002563.

XX 03-FEB-2000; 2000GB-00002564.

XX 20-SEP-2000; 2000GB-00023059.

XX 20-SEP-2000; 2000GB-00023060.

XX (MICR-) MICROSCIENCE LTD.

XX Chatfield SN;

XX WPI; 2001-476280/51.

XX N-PSDB; AAB47030.

XX Novel peptide encoded by a virulence gene of Salmonella typhimurium is
 PT useful for the treatment or prevention of a condition associated with
 PT infection by Salmonella or Gram-negative bacteria, e.g., gastro-
 PT enteritis.

XX Claim 4; Page 39-41; 63pp; English.

XX The invention provides Salmonella typhimurium virulent proteins (AAB85675
 CC -688) and nucleotide sequences (AAH47025-036) encoding the virulent
 CC proteins. The proteins can be expressed by standard recombinant
 CC methodology. The virulent genes and proteins and vaccines and antibodies
 CC raised against the proteins may be used for the manufacture of a
 CC medicament for use in the treatment or prevention of a condition
 CC associated with infection by Salmonella or Gram-negative bacteria, e.g.,
 CC gastro-enteritis. They may also be used for veterinary treatment as well
 CC as in a screening assay for the identification of an antimicrobial drug.
 CC The present sequence represents a Salmonella virulent protein

XX Sequence 297 AA;

XX Query Match 23.5%; Score 272.5; DB 4; Length 297;

XX Best Local Similarity 31.9%; Pred. No. 2.9e-21;

XX Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;

QY 6 KKLQGLVAATITMTENGSENFVIGQYVDLVKQGVKAFVNGTGTGEGLSLSVSRQ 65
 DB 3 KALQGVMAALITPDHQQQLDSESLRLVRFNI--QGGIDGLVGGSTGEAFVQSLAERQ 61

QY 66 VAEFWTKGDKLDQVIVHVGALSLSKESQELAQHAAEIGADGIAVIAPFLKPTWKDILI 125
 DB 62 VLEIVAEAEAKGI-TLIAHVGTVSTAESQQLASAKRYGFDVAVSATVPFFY-PFSFEHC 119

QY 126 NFLKEVAAAPALPFFYYHPIALTGVKIRAEELLDGILDKIPTFGKLFSDTDLRFGQC 185
 DB 120 DHYRAIDSADGLEWVYVNIIPALSGVKLTLDQI--NTLVLPGVGALKQTSGLDYQWEQI 177

QY 186 VTCNRQQQFAFLF-GVDEQLLSALVWGATGAVGS 218

DB 178 --RRHPDLVLYNGYDEIFASGLLAGADGGIGS 208

```

RESULT 15
AAB96342
ID AAB96342 standard; protein; 297 AA.
XX
XX AAB96342;
XX
XX
DT 29-OCT-2001 (first entry)
XX
XX Putative dihydrodipicolinate synthetase/N-acetylneuraminase lyase.
XX
XX Hyperthermophilic archaeon; hyperthermophilic protein.
XX
XX Pyrococcus abyssi.
XX
XX FR2792651-A1.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999; 99FR-00005034.
XX
XX 21-APR-1999; 99FR-00005034.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX (IPRE-) IPREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
XX MPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins
XX useful in industry.
XX
XX Claim 7; Page 1018-1019; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such P. abyssi protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade. Note: This patent is in the same patent family as
XX WO2000065062, which contains additional sequences as shown in AAB99132-
XX AAB99143, AAF75903-AAH75920 and AAG66436
XX
XX Sequence 297 AA:
XX
Query Match 22.0%; Score 254.5; DB 4; Length 297;
Best Local Similarity 30.3%; Pred. No. 2.9e-19;
Matches 66; Conservative 49; Mismatches 82; Indels 2; Gaps 7;

Cy 6 KKLQGLVAALITPTWENGESVIGQVYLVKQGVKNIFVNGTGTGEGSLVSERRQ 65
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 KYMEGVIVPLVTPREDYSDIDFEALWHIRFL-BEKGVHGIFVNSTTGEFTSLNTERKM 60
Qy 66 VAEEW--VTGKGDLDQVLIHWGALSKEOELAAHAEIGADGIVATPZELKPTKDI 123
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 LAEAGRIITG3-----MYLVGTGTSTLEVLTSRHAEDIGAGIVIVAPYCKKDEEI 115
Qy 124 LIINFLEKVAAPALPFYHYHIPALTYGKIRAEELLDIGLDKI-----PTFQGLAFSTDL 179
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 LKHF--SNVAERVDPITVVAIPSC-----ANFVPVDIIRKVSLEYSNIIGVKASVDSL 167
Qy 180 LDFGQCVD-QNRQOQAFPLFGWDEQLLSALVMGATGAV 216
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : :
168 TYLOELIEVKEERKDFVFTGLDQVFLSTLLLGDDGG- 205

```

Search completed: May 6, 2004, 09:04:13
Job time : 38.4251 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: May 6, 2004, 09:00:50 ; Search time 8.7363 Seconds
(without alignments)
2532.427 Million cell updates/sec

Title: US-09-930-440b-2

Perfect score: 1159

Sequence: 1 MAPPKKLGGLVAATTPMT.....GATGAGVGFVSRDLSTLISN 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	443	38.2	297	2 AG0367	probable N-acetylneuraminat
2	430	37.1	298	2 E82157	probable N-acetylneuraminat
3	274.5	23.7	297	1 WZECN	N-acetylneuraminat
4	274.5	23.7	297	2 E85986	N-acetylneuraminat
5	274.5	23.7	297	2 B91141	N-acetylneuraminat
6	272.5	23.5	297	2 AB0908	N-acetylneuraminat
7	256	22.1	305	2 B95195	N-acetylneuraminat
8	252.5	21.6	305	2 B95154	N-acetylneuraminat
9	251.5	21.7	294	2 H75033	dihydrodipicolinat
10	242.5	20.9	293	2 E89796	N-acetylneuraminat
11	241	20.8	313	2 A84203	dihydrodipicolinat
12	240.5	20.8	293	2 G64050	N-acetylneuraminat
13	237.5	20.5	299	2 A92878	probable N-acetylneuraminat
14	232	20.0	302	2 D90489	dihydrodipicolinat
15	231.5	20.0	283	2 A6207	dihydrodipicolinat
16	231.5	20.0	289	2 E90397	dihydrodipicolinat
17	230.5	19.9	287	2 C71135	probable dihydrodipicolinat
18	224.5	19.4	291	2 H81878	probable dihydrodipicolinat
19	222.5	19.2	301	1 C46266	probable dihydrodipicolinat
20	222	19.2	294	2 E70398	dihydrodipicolinat
21	218.5	18.9	319	1 S56523	dihydrodipicolinat
22	218	18.8	309	2 D64752	dihydrodipicolinat
23	217.5	18.8	291	2 D81141	dihydrodipicolinat
24	208	17.9	294	2 AH2265	dihydrodipicolinat
25	208	17.9	302	2 B90485	dihydrodipicolinat
26	205.5	17.7	294	2 AG2702	dihydrodipicolinat
27	205.5	17.7	294	2 G97484	dihydrodipicolinat
28	204.5	17.6	292	2 G83949	dihydrodipicolinat
29	203.5	17.6	290	2 E46665	dihydrodipicolinat

30	201.5	17.4	289	2 E64330	dihydrodipicolinat
31	201	17.3	301	2 S76023	dihydrodipicolinat
32	200	17.3	298	2 A64058	dihydrodipicolinat
33	197.5	17.0	317	2 T36926	dihydrodipicolinat
34	197.5	17.0	333	1 E53308	mosa protein - Rhi
35	196.5	17.0	292	2 C83520	dihydrodipicolinat
36	196.5	17.0	295	2 F83867	dihydrodipicolinat
37	195	16.8	293	2 B97193	dihydrodipicolinat
38	195	16.8	299	2 T35844	dihydrodipicolinat
39	192	16.6	294	2 F87397	dihydrodipicolinat
40	189.5	16.4	307	2 AG3414	dihydrodipicolinat
41	188	16.2	300	2 E70879	dihydrodipicolinat
42	188	16.2	308	2 G90504	2-keto-3-deoxy glu
43	187.5	16.2	294	2 B72246	dihydrodipicolinat
44	186	16.0	289	2 F69363	dihydrodipicolinat
45	185.5	16.0	295	2 C89916	dihydrodipicolinat

ALIGNMENTS

RESULT 1

AG0367

Probable N-acetylneuraminat lyase (EC 4.1.3.3) [imported] - Yersinia pestis (strain CO C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C/Accession: AG0367

R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AG0367

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-297 <KUR>

A/Cross-references: GB:AL590842; PIDN:CAC92266.1; PID:gl5980977; GSPDB:GNC0175

C/Genetics:

A/Gene: YPO3024

C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 38.2%; Score 443; DB 2; Length 297;

Best Local Similarity 43.7%; Pred. No. 4.9e-29;

Matches 93; Conservative 41; Mismatches 75; Indels 4; Gaps 4;

QY 6 KKLQGLVAATITPTMTENGESVIGQYDYLKKEGVKNIFVNGITGESLSVSERRQ 65

Db 2 KKLTHIAAPHTPFDEQGEVNVPEVDIAEHLND-GVKGVYVCGTGEIGHSVDERRK 60

QY 66 VAEWYTKGDKLDQVIIHVGLSLKESQELAQAAEIGADGIAVIAPFLKPTWKDILI 125

Db 61 IAEKRWNAAGKL-SITLHTGALSIRKDAVLSHAETLDFATSAIGCPFFKGNLDELI 119

QY 126 NFKLKEVAAAPALPFYHYHIPALYGVKIRAEELLDGILDKIPTFQGLKESDLDLDFGQC 185

Db 120 AYCOALAAAPSGFYHYH-SGMSGVNLMEQFLKAKESKIPNLSGTFKFNADLYEFQRC 178

QY 186 VQNRQQQAFYFGVDEQLSLVWGCATGAVGS 218

Db 179 L-RVSGKGFDFPFVDEHLPGGLAVGAIGAVGS 210

RESULT 2

E82157

probable N-acetylneuraminat lyase VC1776 [imported] - Vibrio cholerae (strain N16961 s C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: E82157

R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J. chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragci, I.; Sellers, I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: AB2C35; MUID:2046833; PMID:10952301
A:Accession: E82157
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <HEL>
A:Cross-references: GB:AE004255; GB:AE003852; NID:g9656299; PIDN:AAF94925.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1776
A:Map position: 1

Query Match 37.1%; Score 430; DB 2; Length 298;
Best Local Similarity 44.1%; Pred. No. 5.7e-28;
Matches 94; Conservative 39; Mismatches 76; Indels 4; Gaps 4;

Qy 6 KKLQGLVAATITPMTENGELNFSVIGQYVDYLVKEQGVKNI FVNGTTGEGLSLSVSERRQ 65
Db 2 KKLGLIAAPHTPTTKDNKVNFAADQLAELLI EQVGKAVVCGTTGEGIHCSVEERKA 60

Qy 66 VAEWVTGKDKLDQVLIHVGALSLSKESQELAQHAEIGADGIAVIAPFFLKPTWKDI 125
Db 61 IAEWRVKAQDKLD-VILHTGALSTVDITNLTEHAETLDFATSAIGPCFFKPGSGWDLV 119

Qy 126 NLFKEVAAAAPALPFYVHIPALTVGKIRABELLDGILDKIPTFOGLKESDITLLDFQC 185
Db 120 EYCAQVAAAASKGFIYVH-SGMSGVNLJLQFLLKSGQRIPNLXGAFENADLYEYQRC 178

Qy 186 VDQNRQQQFAFLFGVDEQLLSALVMGATGAVGS 218
Db 179 V-RVSNRRKFDIPFGVDEFLPAGLAUGAVGAVGS 210

RESULT 3
WZECN
N-acetylneuraminatase lyase (EC 4.1.1.3) - *Escherichia coli* (strain K-12)
A:Alternate names: N-acetylneuraminatase algolase
C:Species: *Escherichia coli*
C:Date: 31-Mar-1988 #sequence revision 30-Sep-1990 #text_change 01-Mar-2002
C:Accession: JF0002; A24109; C65114
R:Kawakami, B.; Kudo, T.; Narahashi, Y.; Horikoshi, K.
Agric. Biol. Chem. 50, 2155-2159, 1986
A:Title: Nucleotide sequence of the N-acetylneuraminatase lyase gene of *Escherichia coli*.
A:Reference number: A90021
A:Accession: JF0002
A:Molecule type: DNA
A:Residues: 1-297 <KAW>
A:Experimental source: K12, strain H3101
R:Ohta, Y.; Watanabe, K.; Kimura, A.
Nucleic Acids Res. 13, 8843-8852, 1985
A:Title: Complete nucleotide sequence of the *E. coli* N-acetylneuraminatase lyase.
A:Reference number: A93603; MUID:86093682; PMID:3909108
A:Accession: A24109
A:Molecule type: DNA
A:Residues: 1-69, 'G', '71-83', 'T', '85-297' <OHT>
A:Experimental source: K12, strain C500
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C65114
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-297 <BLAT>
A:Cross-references: GB:AE000402; GB:J00096; NID:g1789619; PIDN:AAAC76257.1; PID:g1789620;
A:Experimental source: strain K-12, substrain MG1655
C:Comment: This enzyme catalyzes the conversion of N-acetylneuraminic acid to pyruvate
C:Genetics:
A:Gene: nanA; npl
A:Map position: 69 min
C:Superfamily: dihydrodipicolinate synthase
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

F165/Active site: Lys (covalent pyruvate-binding) #status predicted

Query Match 23.7%; Score 274.5; DB 1; Length 297;
Best Local Similarity 34.3%; Pred. No. 3.7e-15;
Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

Qy 8 LQGLVAATITPMTENGELNFSVIGQYVDYLVKEQGVKNI FVNGTTGEGLSLSVSERRQ-- 65
Db 5 LRGVAAALTPDQQQALDKASRLRVQFNI--QQGIDGLYGGSTGEAFVQSLSREQVL 63

Qy 66 --VAEWTGKDKLDQVLIHVGALSLSKESQELAQHAEIGADGIAVIAPFFLKPTWKDI 123
Db 64 EIVAE--AKGKIL--IAHVGCVSTAESQQLASAKRYGFDASVATPFY--PFSPEE 117

Qy 124 LINFKEVAAAAPALPFYVHIPALTVGKIRABELLDGILDKIPTFOGLKESDITLLDFG 183
Db 118 HCDHYRAIDSADGLPMVWVYNIPALSGVKLTLDQI--NTLVTLPGVGALKQTSGLDYQME 175

Qy 184 QCVQNRQQQFAFLFGVDEQLLSALVMGATGAVGS 218
Db 176 QI---RREHPDLVLYNGYDEIFASGLLAGAGGGIGS 208

RESULT 4
E85986
N-acetylneuraminatase lyase (EC 4.1.1.3) - *Escherichia coli* (strain O157:H7, substrain EI
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: E85986
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 523-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85986
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:Cross-references: GB:AE005174; NID:gl2517843; PIDN:AAOS8353.1; GSPDB:GN00145; UWGP:Z4
A:Experimental source: strain O157:H7, substrain ED1933
C:Genetics:
A:Gene: nanA
C:Superfamily: dihydrodipicolinate synthase
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 23.7%; Score 274.5; DB 2; Length 297;
Best Local Similarity 34.3%; Pred. No. 3.7e-15;
Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

Qy 8 LQGLVAATITPMTENGELNFSVIGQYVDYLVKEQGVKNI FVNGTTGEGLSLSVSERRQ-- 65
Db 5 LRGVAAALTPDQQQALDKASRLRVQFNI--QQGIDGLYGGSTGEAFVQSLSREQVL 63

Qy 66 --VAEWTGKDKLDQVLIHVGALSLSKESQELAQHAEIGADGIAVIAPFFLKPTWKDI 123
Db 64 EIVAE--AKGKIL--IAHVGCVSTAESQQLASAKRYGFDASVATPFY--PFSPEE 117

Qy 124 LINFKEVAAAAPALPFYVHIPALTVGKIRABELLDGILDKIPTFOGLKESDITLLDFG 183
Db 118 HCDHYRAIDSADGLPMVWVYNIPALSGVKLTLDQI--NTLVTLPGVGALKQTSGLDYQME 175

Qy 184 QCVQNRQQQFAFLFGVDEQLLSALVMGATGAVGS 218
Db 176 QI---RREHPDLVLYNGYDEIFASGLLAGAGGGIGS 208

RESULT 5
B91141
N-acetylneuraminatase lyase [imported] - *Escherichia coli* (strain O157:H7, substrain RIME
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence revision 18-Jul-2001 #text_change 02-Aug-2002
C:Accession: B91141
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.-G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands
A:Reference number: A99629; PMID:21156231; PMID:11253796
A:Accession: B91141
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA037521.1; PID:g13363571; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
C:Gene: ECs4098
C:Superfamily: dihydrodipicolinate synthase

Query Match 23.7%; Score 274.5; DB 2; Length 297;
Best Local Similarity 34.3%; Pred. No. 3.7e-15;
Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

QY 6 LQGLVAATITPMTENGSEINFSVIGQYVDYLKVEQGVKNIFVNGTTGEGLSLSVSRQ-- 65
DB LRGVMAALLTPDQQAALDXASRLRVQFNT--QQGIDGLYVGSGTGEAFVQSLSRQVL 63
QY 66 --VAEEWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPEFLKPTWKDIL 123
DB EIVABE--AKGKIL--IAHVGCYSTAESQQLAASAKYGFDAVSATVPFYY-PFSFEH 117
QY 124 LINFLEKVAAPALPFYHYHIPALTGVKIRAEELDGLDKIPFGQLKFSDTLLDFG 183
DB HCDHYRAIDSGDLGFMVYNIPALSGVKLTLDQI--NTLVLPGVGALKQTSGLDYQME 175
QY 184 QCVQDNROQQFAFLR-GVDEQQLSALVMGATGAVS 218
DB QT---RRHPDLVLVNGYDEIFASGLLAGAGGGIGS 208

RESULT 6
AB0908
K-acetylneuraminatase lyase [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhimurium
A:Species: *Salmonella enterica* subsp. *enterica* serovar Typhimurium
A:Note: This species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB0908
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhimurium
A:Reference number: AB0502; PMID:21534947; PMID:11677608
A:Accession: AB0908
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07856.1; PID:g16504403; GSPDB:GN00176
C:Genetics:
A:Gene: STX3520
C:Superfamily: dihydrodipicolinate synthase

Query Match 23.5%; Score 272.5; DB 2; Length 297;
Best Local Similarity 31.9%; Pred. No. 5.3e-15;
Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;

QY 6 KKLQGLVAATITPMTENGSEINFSVIGQYVDYLKVEQGVKNIFVNGTTGEGLSLSVSRQ 65
DB 3 KALQGLVAALLTPDQQAALDXASRLRVQFNT--QQGIDGLYVGSGTGEAFVQSLSRQ 61
QY 66 VAEWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPEFLKPTWKDIL 125
DB 62 VLEIVAEAKGI-TLIAHVGVSTAESQQLAASAKRYGFDAVSATVPFYY-PFSFEHC 119
QY 126 NLFKEVAAPALPFYHYHIPALTGVKIRAEELDGLDKIPFGQLKFSDTLLDFGQC 185
DB 120 DHYRAIDSGDLGFMVYNIPALSGVKLTLDQI--NTLVLPGVNALKQTSGLDYQME 177

QY 186 VQDNROQQFAFLR-GVDEQQLSALVMGATGAVS 218
DB 178 --RRHPDLVLVNGYDEIFASGLLAGAGGGIGS 208
RESULT 7
B95195
K-acetylneuraminatase lyase, probable [imported] - *Streptococcus pneumoniae* (strain TIGR4)
A:Species: *Streptococcus pneumoniae*
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: B95195
A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.
A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: B95195
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75755.1; PID:g14973168; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI676

Query Match 22.1%; Score 256; DB 2; Length 305;
Best Local Similarity 31.2%; Pred. No. 1.3e-13;
Matches 67; Conservative 42; Mismatches 94; Indels 12; Gaps 6;

QY 6 KKLQGLVAATITPMTENGSEINFSVIGQYVDYLKVEQGVKNIFVNGTTGEGLSLSVSRQ 65
DB 5 KKEGVIPAFYACVDEQGEVSPERTRALVQYFI-DKGQGLYVNGSGECITQGVSDRKL 63
QY 66 VAEWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPEFLKPTWKDIL 125
DB 64 LLEEVNVAAGKL-TLIAHVACNNTKDSNLAHSLGVDALATIPPYIFR-LPYSVA 121
QY 126 NLFKEVAAPALPFYHYHIPALTGVKIRAE--ELLGDKIPFGQLKFSDTLLDF 182
DB 122 KYWNDISSAAPTVDYVNIPLQAGVALTPSLYTEML-----KNPEVIGKSSMPVQDI 176
QY 183 GCVQDNROQQFAFLR-GVDEQQLSALVMGATGAVS 217
DB 177 QTFVSLGGEDHIVF-NGPDEQFLGGLMGARAGIG 210

RESULT 8
B95154
K-acetylneuraminatase lyase [imported] - *Streptococcus pneumoniae* (strain TIGR4)
A:Species: *Streptococcus pneumoniae*
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: B95154
A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.
A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: B95154
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75427.1; PID:g14972811; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI329

Query Match 21.8%; Score 252.5; DB 2; Length 305;
Best Local Similarity 31.1%; Pred. No. 2.4e-13;


```
RESULT 12
G64050
N-acetylneuraminase lyase (EC 4.1.3.3) [similarity] - Haemophilus influenzae (strain Rd
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Jul-2000
C:Accession: G64050
R.Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.;
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A.:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7342800
A:Accession: G64050
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-293 <TIGR>
A:Cross-references: GB:U32700; GB:L42023; NID:g3212181; PIDN:AAC2181.4.1; PID:g1573098; T
C:Superfamily: dihydrodipicolinate synthase
C:Keywords: carbon-carbon lyase; oxo-acid-lyase
F:164/Active site: Lys (covalent pyruvate-binding) #status predicted
```

```
Query Match 20.8%; Score 240.5; DB 2; Length 293;
Best Local Similarity 28.9%; Pred. No. 2.3e-12;
Matches 66; Conservative 44; Mismatches 81; Indels 37; Gaps 6;

QY 6 KKLGLVAATITPMTENGSEINFSVIGQYDYLKVEQGVKNIFVNGTTGEGLSLSVSERRQ 65
DB 2 RDLGIFSAALLVSFNEGDTINEKGLAQIIRHNDKMKVGLYGGSTGENFPM-STBEKE 61
QY 66 VAEWVTGKDKLDQVLIHVGLSLSKESQELAQHAAEIGADGIAVIAPPLKWTXIDILI 125
DB 62 IFRIAKDEAKQI-ALLAQGVSNLKEAVLSKATLGVDCLSAVTPFYK----- 112
QY 126 NFKKEVAAAAPALPFY-----YHIALTGKVKIRAEELLDGILDKIPTFQGL 172
DB 113 -----FSPPEIKHYVDITIAETGNMIVTSIPFLTGVMNGIEQF--GELYKNPKVLGV 163
QY 173 KPSDITD--ILLDQGVQDQNRQQQFAELFGVDEQLLSALVMGATGAVGS 218
DB 164 KFTAGDYLLERLKKAYPNH---LIWAGFDEWMLPAASLGVGDGATGS 207
```

```
RESULT 13
A95878
probable N-acetylneuraminase lyase subunit protein [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 27-Nov-2001
C:Accession: A95878
R.Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A95878
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC48689.1; PID:g15140161; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R.Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Rubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, B.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leilaure,
Hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: nanA; SWB20299
A:Genome: plasmid
C:Superfamily: dihydrodipicolinate synthase
```

```
Query Match 20.5%; Score 237.5; DB 2; Length 299;
Best Local Similarity 29.8%; Pred. No. 4.1e-12;
Matches 62; Conservative 49; Mismatches 92; Indels 9; Gaps 7;

QY 7 KKLGLVAATITPMTENGSEINFSVIGQYDYLKVEQGVKNIFVNGTTGEGLSLSVSERRQ 66
DB 2 KLEGIYSALLTSPSEESIDRQALGALVDQVR-LGIDGVYVGGSSGEAMLOSILDERADY 60
QY 67 ABEWVTGKDKLDQVLIHVGLSLSKESQELAQHAAEIGADGIAVIAPPLKWTXIDILI 126
DB 61 LSDVAAAASGRLL-TLIAHVGTIATRDALRLSQHAAKSGYQAISAPFPFYDFSRPEVMAH 119
QY 127 FLKEVAAAAPALPFYHYHIALTGKVKIRAEELLDGILDKIPTFQGLKFSDTLLDFQCV 186
DB 120 Y-RELADWS-ALPLIVYNFFARTS-QFTLPFLVE--LLSHENIIGIKHTSDMFQLERI- 173
QY 187 DQNRQQQFAELFGVDEQLLSALVMGATGAVGS 218
DB 174 -RHAVPDALVINGYDEMCLAGFANGAGGAGT 204
```

```
RESULT 14
D90489
dihydrodipicolinate synthase (dapA-3) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: D90489
R.;She, Q.; Singh, R.K.; Confalonieri, P.; Zivanovic, Y.; Allard, G.; Awavez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, I.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: D90489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <KUR>
A:Cross-references: GB:AE006641; NID:g13816479; PIDN:NAK43171.1; GSPDB:GN00155
C:Genetics:
A:Gene: dapA-3
```

```
Query Match 20.0%; Score 232; DB 2; Length 302;
Best Local Similarity 31.8%; Pred. No. 1.2e-11;
Matches 68; Conservative 40; Mismatches 96; Indels 10; Gaps 6;

QY 7 KKLGLVAATITPMTENGSEINFSVIGQYDYLKVEQGVKNIFVNGTTGEGLSLSVSERRQ 66
DB 2 RLEGVVVPLITPLEDYSIDKEGLKWLVSYL-SENGVWGIIVNSVSGEFASLTIDEMKLV 60
QY 67 ABEWVTGKDKLDQVLIHVGLS--LKESQELAQHAAEIGADGIAVIAPPLKWTXIDIL 124
DB 61 TK--IALDSRKSYSTLIIFAGASTNFTETIKLAKPFOMGVDAIVVMAPYFFKVRKELL 118
QY 125 INFLKEVAAAAPALPFYHYHIALTGKVKIRAEELLDGILDKIPTFQGLKFSDTLLDFGQ 184
DB 119 DHP--SMIAEKVDLPILIIYNIPLFTGIDIPV-SVKMLVSYQHSNIIGTKVTLDSLFFKK 175
QY 185 CVDQNRQ--QQFAFLFGVDEQLLSALVMGATGAV 216
DB 176 LISEIREIKRDFSILTGDFEYLLPLLYMGNGGV 209
```

```
RESULT 15
A69207
dihydrodipicolinate synthase (EC 4.2.1.52) MTH801 [similarity] - Methanobacterium therm
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 31-Mar-2000
C:Accession: A69207
R.;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
```

Search completed: May 6, 2004, 09:09:41
Job time : 10.9863 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:56:04 ; Search time 6.04821 Seconds
(without alignments)
1980.112 Million cell updates/sec

Title: US-09-930-440B-2

Perfect score: 1.59

Sequence: 1 YAFPRKKLQGLVAATITPT.....GATGAGSFVSRDLSTLLSN 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274.5	23.7	296	1 NANA_ECOLI	P06995 escherichia
2	273.5	23.6	296	1 NANA_ECOLI	Q8fd58 escherichia
3	272.5	23.5	297	1 NANA_SALTI	Q8z3f0 salmonella
4	272.5	23.5	297	1 NANA_SALTI	Q8z3f0 salmonella
5	247.5	21.4	293	1 NANA_PASMU	Q9ckb0 pasteurella
6	246.5	21.3	305	1 NANA_ECOLI	Q8fdi7 escherichia
7	242.5	20.9	293	1 NANA_STAAW	Q99wt1 staphylococ
8	242.5	20.9	293	1 NANA_STAAW	Q8nyc7 staphylococ
9	241.5	20.8	292	1 NANA_LACPL	P59407 lactobacill
10	241.5	20.8	304	1 NANA_HALNI	Q9hs19 halobacteri
11	240.5	20.8	293	1 NANA_HAEIN	P44539 haemophilus
12	237.5	20.5	299	1 NANA_RHIME	Q92wp0 rhizobium m
13	234.5	20.2	300	1 DAPA_METKA	Q8tu24 methanopyru
14	231.5	20.0	283	1 DAPA_METH	Q26892 methanobact
15	226	19.5	291	1 DAPA_METMA	Q8px17 methanosarc
16	225.5	19.5	290	1 NANA_FUSNN	Q8rdn6 fusobacteri
17	224.5	19.4	291	1 DAPA_NEIMA	Q9jnu9 neisseria m
18	222.5	19.2	301	1 DAPA_CORGL	P19808 corynebacte
19	222	19.2	294	1 DAPA_AQUA	O67216 aquifex aeo
20	220.5	19.0	297	1 DAPA_THETN	Q8rb15 thermoaer
21	218.5	18.9	288	1 NANA_CLOPE	Q8s4k9 clostridium
22	218.5	18.9	301	1 YGHH_ECOLI	P39359 escherichia
23	218	18.8	309	1 YAGE_ECOLI	P75682 escherichia
24	217.5	18.8	291	1 DAPA_NEIMB	Q9jzr4 neisseria m
25	216.5	18.7	291	1 DAPA_METAC	Q8thp1 methanosarc
26	216	18.6	302	1 DAPA_PROMA	P49423 prochloroco
27	213.5	18.4	301	1 DAPA_COREF	Q8rqm8 corynebacte
28	213	18.4	295	1 DAPA_COREH	Q8eqj1 oceanobacil
29	208	17.9	294	1 DAPA_ANASP	Q9qyvl anabaena sp
30	207.5	17.9	294	1 DAPA_STAEP	Q8cp96 staphylococ
31	206	17.8	291	1 DAPA_CLOPE	Q8xj56 clostridium
32	206	17.8	307	1 DAPA_LEPIN	Q8f132 leptospira
33	205.5	17.7	294	1 DAPA_AGRTS	Q8ug13 agrobacteri

34	204.5	17.6	292	1 DAP2_BACHD	Q9ka91 bacillus ha
35	203.5	17.6	290	1 DAP2_BACSU	Q04796 bacillus su
36	202.5	17.5	294	1 DAP1_RHIME	Q92x55 rhizobium m
37	202.5	17.5	294	1 DAP2_RALSO	Q8y099 ralsconia s
38	202	17.4	296	1 DAPA_BRAJA	Q9rh76 bradyrhizob
39	201.5	17.4	289	1 DAPA_METJA	Q57695 methanococc
40	201	17.3	301	1 DAPA_VSNY3	Q55513 syntechocyst
41	200	17.3	298	1 DAPA_HAEIN	P43797 haemophilus
42	199.5	17.2	301	1 DAPA_BIFLO	Q8527 bifidobacte
43	199	17.2	296	1 DAPA_SYNEL	Q8d1k4 syntecococc
44	197.5	17.0	305	1 DAP2_STRCO	Q9x9w0 streptomyce
45	197.5	17.0	333	1 MOSA_RHIME	Q07607 rhizobium m

ALIGNMENTS

RESULT 1	ID	NANA_ECOLI	STANDARD;	PRT;	296 AA.
AC	P06995;				
CT	01-APR-1988 (Rel. 07, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	N-acetylneuraminase lyase (EC 4.1.3.3) (N-acetylneuraminic acid aldolase) (N-acetylneuraminase pyruvate-lyase) (Sialic acid lyase)				
DE	(Sialate lyase) (Sialic acid aldolase) (NALase).				
GN	NANA OR NPL OR B3225 OR Z4583 OR ECS4098 OR SF3261 OR S3478.				
OS	Escherichia coli,				
OS	Escherichia coli O157:H7, and				
OS	Shigella flexneri.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562, 83334, 623;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	SPECIES=E.coli; STRAIN=K12;				
RX	MEDLINE=86093682; PubMed=3909108;				
RA	Ohta Y., Watanabe K., Kimura A.;				
RT	"Complete nucleotide sequence of the E. coli N-acetylneuraminase lyase."				
RL	Nucleic Acids Res. 13:8843-8852 (1985).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=E.coli; STRAIN=JEL1011;				
RX	Kawakami B., Kudo T., Narahashi Y., Horikoshi K.;				
RA	"Nucleotide sequence of the N-acetylneuraminase lyase gene of Escherichia coli."				
RT	Agric. Biol. Chem. 50:2155-2158 (1986).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=E.coli; STRAIN=K12 / MG1655;				
RX	MEDLINE=97426617; PubMed=9278503;				
RA	Blattner E.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;				
RT	"The complete genome sequence of Escherichia coli K-12."				
RL	Science 277:1453-1474 (1997).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;				
RX	MEDLINE=21074935; PubMed=11206551;				
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;				
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."				
RL	Nature 409:529-533 (2001).				
RN	[5]				
RP	SEQUENCE FROM N.A.				

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMEL; X03345; CAA27051.1; -
 CC EMEL; D00067; BAA00046.1; -
 CC EMEL; U18997; AAA58027.1; -
 CC EMEL; AE000402; AAC76257.1; -
 CC EMEL; AE000550; AAG59353.1; -
 CC EMEL; AF002564; BAB37521.1; -
 CC EMEL; AE015337; AAN44725.1; -
 CC EMEL; AE016989; AAP18538.1; -
 CC PIR; B91141; B91141.
 CC PIR; E85986; E85986.
 CC PIR; JP0002; WZECN.
 CC PDB; 1NAL; 15-SEP-95.
 CC PDB; 1FDY; 22-OCT-97.
 CC PDB; 1FDZ; 22-OCT-97.
 CC ECO2DBASE; G032.6; 6TH EDITION.
 CC EcoGene; EG10637; nanA.
 CC HAMAP; MF_01237; -; 1.
 CC InterPro; IPR002220; DHDES.
 CC InterPro; IPR005264; NanA.
 CC Pfam; PF00701; DHDES; 1.
 CC PRINTS; PR00146; DHPICNTHASE.
 CC ProDom; PD001859; DHDES; 1.
 CC TIGRfam; TIGR00683; nanA; 1.
 CC PROSITE; PS00665; DHDES; 1.
 CC PROSITE; PS00666; DHDES_2; 1.
 CC Carbohydrate metabolism; Lyase; Schiff base; 3D-structure;
 CC Complete proteome.
 CC INIT_MET 0
 CC ACT_SITE 136 136 INVOLVED IN PROTON TRANSFER DURING
 CC ACT_SITE 164 164 CLEAVAGE.
 CC ACT_SITE 69 69 SCHIFF-BASE WITH N-ACETYLNEURAMINATE.
 CC ACT_SITE 83 83 A -> G (IN REF. 1).
 CC ACT_SITE 281 281 S -> T (IN REF. 1).
 CC ACT_SITE 7 10 L -> Q (IN REF. 1).
 CC ACT_SITE 15 15
 CC ACT_SITE 17 18
 CC ACT_SITE 21 21
 CC ACT_SITE 23 35
 CC ACT_SITE 36 37
 CC ACT_SITE 40 43
 CC ACT_SITE 46 49
 CC ACT_SITE 50 52
 CC ACT_SITE 55 69
 CC ACT_SITE 70 72
 CC ACT_SITE 74 78
 CC ACT_SITE 84 96
 CC ACT_SITE 97 98
 CC ACT_SITE 101 105
 CC ACT_SITE 114 127
 CC ACT_SITE 128 130
 CC ACT_SITE 133 137
 CC ACT_SITE 139 142
 CC ACT_SITE 148 154
 CC ACT_SITE 155 156
 CC ACT_SITE 158 159
 CC ACT_SITE 160 165
 CC ACT_SITE 170 179
 CC ACT_SITE 181 182
 CC ACT_SITE 184 187
 CC ACT_SITE 190 192
 CC ACT_SITE 193 198
 CC ACT_SITE 199 200
 CC ACT_SITE 203 206
 CC ACT_SITE 207 208

CC SPECIES=E.coli; STRAIN=0157:H7 / RIMD 0509952;
 CC MEDLINE=21156231; PubMed=11258796;
 CC Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 CC Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka Y., Tobe T.,
 CC Iida T., Takami H., Honda T., Sasakawa C., Ogasawara K., Yasunaga T.,
 CC Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 CC "Complete genome sequence of enterohemorrhagic Escherichia coli
 CC O157:H7 and genomic comparison with a laboratory strain K-12.";
 CC DNA Res. 8:11-22(2001).
 CC [6]
 CC SEQUENCE FROM N.A.
 CC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 CC MEDLINE=2272406; PubMed=12384590;
 CC Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 CC Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 CC Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 CC Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Heu Y.,
 CC Ya J.;
 CC "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 CC through comparison with genomes of Escherichia coli K12 and O157.";
 CC Nucleic Acids Res. 30:4432-4441(2002).
 CC [7]
 CC SEQUENCE FROM N.A.
 CC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
 CC MEDLINE=22590274; PubMed=12704152;
 CC Wei J., Goldberg M.B., Burland V., Venkatesan V.M., Deng W.,
 CC Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 CC Mau B., Perra N.T., Payne S.M., Ruyten-Janecky L.J., Zhou S.,
 CC Schwartz D.C., Blattner F.R.;
 CC "Complete genome sequence and comparative genomics of Shigella
 CC flexneri serotype 2a strain 2457T.";
 CC Infect. Immun. 71:2775-2786(2003).
 CC [8]
 CC SEQUENCE OF 1-23, AND CHARACTERIZATION.
 CC SPECIES=E.coli; STRAIN=K12 / C600 / SF8;
 CC MEDLINE=91264813; PubMed=1646603;
 CC Aisaka K., Igarashi A., Yamaguchi K., Uwajima T.;
 CC "Purification, crystallization and characterization of
 CC N-acetylneuraminase from Escherichia coli.";
 CC Biochem. J. 276:541-546(1991).
 CC [9]
 CC X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 CC SPECIES=E.coli;
 CC MEDLINE=94363237; PubMed=8081752;
 CC Izard T., Lawrence M.C., Malby R.L., Lilley G.G., Colman P.M.;
 CC "The three-dimensional structure of N-acetylneuraminase from
 CC Escherichia coli.";
 CC Structure 2:361-369(1994).
 CC [10]
 CC X-RAY CRYSTALLOGRAPHY (2.45 ANGSTROMS) OF COMPLEXES WITH
 CC HYDROXYPYRUVATE AND PYRUVATE.
 CC SPECIES=E.coli;
 CC MEDLINE=97193395; PubMed=9047371;
 CC Lawrence M.C., Barbosa J.A., Smith B.J., Hall N.E., Pilling P.A.,
 CC Ooi H.C., Marcuccio S.M.;
 CC "Structure and mechanism of a sub-family of enzymes related to N-
 CC acetylneuraminase lyase.";
 CC J. Mol. Biol. 266:381-399(1997).
 CC !- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
 CC (sialic acid) to form pyruvate and N-acetylmannosamine via a
 CC Schiff base intermediate.
 CC !- CATALYTIC ACTIVITY: N-acetylneuraminase = N-acetyl-D-mannosamine +
 CC pyruvate.
 CC !- ENZYME REGULATION: Inhibited by reduction with NaBH(4), and by
 CC Cu(2+) ions, p-chloromercuribenzoate and N-bromosuccinimide.
 CC !- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
 CC !- SUBUNIT: Homotetramer.
 CC !- SUBCELLULAR LOCATION: Cytoplasmic.
 CC !- INDUCTION: By N-acetylneuraminase.
 CC !- MISCELLANEOUS: Optimum temperature for activity is 60 degrees
 CC Celsius and optimum pH is 6.5-7.0.
 CC !- SIMILARITY: Belongs to the DHDES family. NanA subfamily.
 CC -----


```

CC      Schiff base intermediate (By similarity).
CC      -!- CATALYTIC ACTIVITY: N-acetylneuraminase = N-acetyl-D-mannosamine +
CC      Pyruvate.
CC      -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
CC      -!- SUBUNIT: Homotrimer (By similarity).
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC      -!- SIMILARITY: Belongs to the DHDPS family. Nana subfamily.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; AL623278; CAD07856.1; -
CC      EMBL; ABC16845; AAO0793.1; -
CC      HAMAP; MF_02237; -; 1.
CC      InterPro; IPR002220; DHDPS.
CC      InterPro; IPR005264; Nana.
CC      Pfam; PF00701; DHDPS.1.
CC      PRINTS; PS00146; DHPICSTHASE.
CC      ProDom; PDC01859; DHDPS.1.
CC      TIGRFAMs; TIGR00683; nana; 1.
CC      PROSITE; PS00665; DHDPS.1; 1.
CC      PROSITE; PS00666; DHDPS.2; 1.
CC      Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
KW      ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING
FT      ACT_SITE 165 165 CLEAVAGE (BY SIMILARITY).
FT      ACT_SITE 165 165 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
FT      ACT_SITE 165 165 (BY SIMILARITY).
SQ      SEQUENCE 297 AA; 32482 MW; DFFB0F1E4895546 CRC64;
Query Match 23.5%; Score 272.5; DB 1; Length 297;
Best Local Similarity 31.9%; Pred. No. 1.1e-14;
Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;
QY 6 KKLQGLVAATIPPMTEENGSEINFVIGQYVDYLVKQGVKNIFVNGTGGSLSVSRQ 65
Db 3 KALQGVMAALLTFPHQOQDSESLRLVRFNI-QGIDGLVVGSGTGEAFVQSLAEREQ 61
QY 66 VAEWVTYKQKLDQVTHVGLSLKESQELACHAAEIGADGIAVIAPFLKPTWKDILI 125
Db 62 VLEIVAREAKGKI-TLIAHVGTVSTAESQQLASAKRYGFDVAVTPTFY-PFSEEH 119
QY 126 NFLKEVAAAPALPFYVYHIPALTGVKIRABELLQILDKIPTFQGLKFSDTDLDFGQC 185
Db 120 DHYRAILDSADGLPMWVYVNIIPALSGVKLTLDQI--NTLVTLPGVNAIKQTSGLDFQMEQI 177
QY 186 VQNRQQQAFILFGVDEQQLLSALVMGATGAVS 218
Db 178 --RRAHPDLVLYNGYDEIFASGLLAGAGGGIGS 208
RESULT 4
NANA SALT STANDARD; PRT; 297 AA.
AC 28ZLQ6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE N-acetylneuraminase lyase (EC 4.1.3.3) (N-acetylneuraminic acid
DE aldolase) (N-acetylneuraminase pyruvate-lyase) (Sialic acid lyase)
DE (Sialate lyase) (Sialic acid aldolase).
GN NANA OR STM3339.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.B.; Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvane E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
CC (sialic acid) to form pyruvate and N-acetylmannosamine via a
CC Schiff base intermediate (By similarity).
CC -!- CATALYTIC ACTIVITY: N-acetylneuraminase = N-acetyl-D-mannosamine +
CC Pyruvate.
CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHDPS family. Nana subfamily.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; AB008854; AAL22208.1; -
CC      HAMAP; MF_01237; -; 1.
CC      InterPro; IPR002220; DHDPS.
CC      InterPro; IPR005264; Nana.
CC      Pfam; PF00701; DHDPS.1.
CC      PRINTS; PR00146; DHPICSTHASE.
CC      ProDom; PDC01859; DHDPS.1.
CC      TIGRFAMs; TIGR00683; nana; 1.
CC      PROSITE; PS00665; DHDPS.1; 1.
CC      PROSITE; PS00666; DHDPS.2; 1.
CC      Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
KW      ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING
FT      ACT_SITE 165 165 CLEAVAGE (BY SIMILARITY).
FT      ACT_SITE 165 165 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
FT      ACT_SITE 165 165 (BY SIMILARITY).
SQ      SEQUENCE 297 AA; 32455 MW; 78D599F104937C4D CRC64;
Query Match 23.5%; Score 272.5; DB 1; Length 297;
Best Local Similarity 31.9%; Pred. No. 1.1e-14;
Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;
QY 6 KKLQGLVAATIPPMTEENGSEINFVIGQYVDYLVKQGVKNIFVNGTGGSLSVSRQ 65
Db 3 KALQGVMAALLTFPHQOQDSESLRLVRFNI-QGIDGLVVGSGTGEAFVQSLAEREQ 61
QY 66 VAEWVTYKQKLDQVTHVGLSLKESQELACHAAEIGADGIAVIAPFLKPTWKDILI 125
Db 62 VLEIVAREAKGKI-TLIAHVGTVSTAESQQLASAKRYGFDVAVTPTFY-PFSEEH 119
QY 126 NFLKEVAAAPALPFYVYHIPALTGVKIRABELLQILDKIPTFQGLKFSDTDLDFGQC 185
Db 120 DHYRAILDSADGLPMWVYVNIIPALSGVKLTLDQI--NTLVTLPGVNAIKQTSGLDFQMEQI 177
QY 186 VQNRQQQAFILFGVDEQQLLSALVMGATGAVS 218
Db 178 --RRAHPDLVLYNGYDEIFASGLLAGAGGGIGS 208
RESULT 5
NANA PASMU STANDARD; PRT; 293 AA.
AC Q9CKH0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```


DB 122 Y-RKIDSADGLPMVYVYNIPALSGVRFSLDQINE--LVTIPRVCAKQTSGLFQWBOI- 177

QY 187 QNRQOQAFILFGVDEQLLSALVMGATGAVGS 218

DB 178 -KRNHPELVLYNGYDEIFASGLIAGADGGIGS 208

RESULT 7

NANA STAAW
ID NANA STAAW STANDARD; PRT; 293 AA.
AC Q99WR1.
DT 13-OCT-2003 (Rel. 42, Created)
DT 13-OCT-2003 (Rel. 42, Last sequence update)
DT 13-OCT-2003 (Rel. 42, Last annotation update)
DE N-acetylneuraminidase (EC 4.1.3.3) (N-acetylneuraminic acid
aldehyde) (N-acetylneuraminidase pyruvate-lyase) (Sialic acid lyase)
DE (Sialate lyase) (Sialic acid aldolase).
GN NANA OR SAV0315 OR SA3304.
OS Staphylococcus aureus (strain Mu50 / ATCC 70669), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 70669, and N315;
RX MEDLINE=21311952; PubMed=1418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Igarashi J., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kankihisa M., Yamashita A., Oshida K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus";
RL Lancet 357:1225-1240(2001).
CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
(Sialic acid) to form pyruvate and N-acetylmannosamine via a
Schiff base intermediate (By similarity).
CC -!- CATALYTIC ACTIVITY: N-acetylneuraminidase = N-acetyl-D-mannosamine +
pyruvate.
CC -!- PATHWAY: N-acetylneuraminic acid (Sialic acid) utilization.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHPS family. Nana subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003359; BAB56477.1; -;
CC EMBL; AP003130; BAB41528.1; -;
CC PIR; E89796; E89796.
CC HSP; P06995; INAL.
CC SWISS-2DPAGE; Q99WR1; STAAW.
CC HAMAP; MF 01237; -; 1.
CC InterPro; IPR002220; DHPS.
CC Pfam; PF00701; DHPS; 1.
CC PRINTS; PR00146; DHPICNTASE.
CC PROSITE; PD001859; DHPS; 1.
CC PROSITE; PS00665; DHPS; 1; FALSE NEG.
CC PROSITE; PS00666; DHPS; 2; FALSE NEG.
KW Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
FT ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING
FT ACT_SITE 165 165 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
FT ACT_SITE 165 165 (BY SIMILARITY).
FT SEQUENCE 293 AA; 33100 MW; 4D07DEFFDCA701 CRC64;

Query Match 20.9%; Score 242.5; DB 1; Length 293;
Best Local Similarity 30.0%; Pred. No. 2.7e-12;
Matches 68; Conservative 46; Mismatches 78; Indels 35; Gaps 8;

QY 6 KKLQGLVAATITETENGESINFSVIGQYVDYLAVKEQGVKNIEVNGTGGELSLVSVERPQ 65

DB 3 KDLKGLYALIVFDENGQVNEQGLKQTAQNAISTEELDLTYNGSSGENFLINTEQKQK 62

QY 66 ---VAEEWVTGKDKLDQVVIHVGLSLKESQELAQAHAIEIGADGLAVIAPFFLKPTWKD 122

DB 63 VFKVAKEAV---GDKV-KLIAQVGSLSLNEATELCKYATELGVDAISALSVTPFYV-PEETE 117

QY 123 ILINELKEVAAPALPFYHYHIPALTGVKIRAE-----ELLDGLDKITFP---QG 171

DB 118 EIRDFYDIIIEATQN-NMIIYAIPLDTGVNIDIEQFSELFNEHKIVGVKYTAPNFFLLER 176

QY 172 LKFSPTDILLDFGQVDQNRQQQAFILFGVDEQLLSALVMGATGAVGS 218

DB 177 IRKAFEDKL-----ILSGFDEMLVQATISGVDAIGS 208

RESULT 8

NANA STAAW
ID NANA STAAW STANDARD; PRT; 293 AA.
AC Q99WR1.
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE N-acetylneuraminidase (EC 4.1.3.3) (N-acetylneuraminic acid
aldehyde) (N-acetylneuraminidase pyruvate-lyase) (Sialic acid lyase)
DE (Sialate lyase) (Sialic acid aldolase).
GN NANA OR MW0292.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA";
RL Lancet 359:1819-1827(2002).
CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
(Sialic acid) to form pyruvate and N-acetylmannosamine via a
Schiff base intermediate (By similarity).
CC -!- CATALYTIC ACTIVITY: N-acetylneuraminidase = N-acetyl-D-mannosamine +
pyruvate.
CC -!- PATHWAY: N-acetylneuraminic acid (Sialic acid) utilization.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHPS family. Nana subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF004823; BAB94157.1; -;
CC HAMAP; MF 01237; -; 1.
CC InterPro; IPR002220; DHPS.
CC Pfam; PF00701; DHPS; 1.
CC PRINTS; PR00146; DHPICNTASE.
CC PROSITE; PD001859; DHPS; 1.
CC PROSITE; PS00665; DHPS; 1; FALSE NEG.
CC PROSITE; PS00666; DHPS; 2; FALSE NEG.
KW Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
FT ACT_SITE 137 137 INVOLVED IN PROTON TRANSFER DURING

FT ACT_SITE 165 165 CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 165 165 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
 FT ACT_SITE 165 165 (BY SIMILARITY).
 SQ SEQUENCE 293 AA; 33042 MW; 62D7DEFFCDAF431 CRC64;
 Query Match 20.9%; Score 242.5; DB 1; Length 293;
 Best Local Similarity 30.0%; Pred. No. 2.7e-12;
 Matches 68; Conservative 46; Mismatches 78; Indels 35; Gaps 8;
 QY 6 KKLQGLVAATITPMTEINGENFVSIGQVYDYLKQKQVKNIFVNGTGTGEGSLSVSRERQ 65
 DB 3 KDLKGLYAALLVPDENGQVNEQLKQIAQNAITEELDGLYNGSSGFLLNTEQKKQ 62
 QY 66 ---VAEWWTKGDKLDQVIVHVCALSKESQELAAHAEIGADGIAVIAPFFKLPWKD 122
 DB 63 VFKVAKAV---GDKV-KLIAQVGLDLNEAIELGKATYELGYDALSAAVTFYV-PTFE 117
 QY 123 ILINFLKEVAAAAPALFFYYVHIIHALTGKIRAS-----ELDGLDKIPTF---QG 171
 DB 118 EIRDYFEIIEATQN-NMIIAIPDLTGWNISIQFSELFNHEKIVGVKTAPEAFLLER 176
 QY 172 LKPSDTLLDFGQCVDQNRQQQAFLEFGVDEQLLSALVMGATGAVGS 218
 DB 177 IRKAFPKL-----ILSGDEMLVQATISGVGGAIGS 208
 RESULT 9
 NANA LACPL STANDARD; PRT; 292 AA.
 ID NANA LACPL
 AC P59407;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE N-acetylneuraminatase (EC 4.1.3.3) (N-acetylneuraminic acid
 aldolase) (N-acetylneuraminatase pyruvate-lyase) (Sialic acid lyase)
 DE (Sialate lyase) (Sialic acid aldolase).
 GN NANA OR LP 3568.
 OS Lactobacillus plantarum.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB 8826 / WCPS1;
 RX MEDLINE=22480296; PubMed=12566566;
 RA Kleerebezem N., Boekhorst J., van Kranenburg R., Molenaar D.,
 RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink K.W.,
 RA Fiers W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
 RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
 RA De Vos W.M., Siezen R.J.;
 RT "Complete genome sequence of Lactobacillus plantarum WCPS1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
 CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
 (sialic acid) to form pyruvate and N-acetylmannosamine via a
 Schiff base intermediate (By similarity).
 CC -!- CATALYTIC ACTIVITY: N-acetylneuraminatase = N-acetyl-D-mannosamine +
 pyruvate.
 CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
 CC -!- SUBUNIT: Homotrimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the DHDPS family. Nana subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: ALG95262; CAJ65647.1; -
 CC HAMAP: MF 01237; -; 1.
 CC InterPro: IPR002220; DHDPS.

DR PFam: PF09701; DHDPS: 1.
 DR PRINTS: PR00146; DHDPS: 1.
 DR PROSITE: PS00665; DHDPS: 1; 1.
 DR PROSITE: PS00666; DHDPS: 2; FALSE NEG.
 KW Carbohydrate metabolism; Lyase; Schiff base; Complete proteome.
 FT ACT_SITE 135 135 INVOLVED IN PROTON TRANSFER DURING
 FT ACT_SITE 135 135 CLEAVAGE (BY SIMILARITY).
 FT ACT_SITE 163 163 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
 FT ACT_SITE 163 163 (BY SIMILARITY).
 SQ SEQUENCE 292 AA; 32708 MW; 4184967AE9C91747 CRC64;
 Query Match 20.8%; Score 241.5; DB 1; Length 292;
 Best Local Similarity 31.9%; Pred. No. 3.3e-12;
 Matches 69; Conservative 38; Mismatches 94; Indels 15; Gaps 6;
 QY 6 KKLQGLVAATITPMTEINGENFVSIGQVYDYLKQKQVKNIFVNGTGTGEGSLSVSRERQ 65
 DB 3 KKL--LYAAQMTAFDKGNINLDGIRALVRNIDVKNVQGLYVCGSTGEAFMLNTEKKQ 60
 QY 66 VAEWWTKGDKLDQVIVHVCALSKESQELAAHAEIGADGIAVIAPF---ELKPTKD 122
 DB 61 VMETVYDEANGAID-LVAQVGLSLNKEAKELAKATDGLGPKLSAVTFYNYNFTPEQND 119
 QY 123 ILINFLKEVAAAAPALFFYYVHIIHALTGKIRABELLDGLDKIPTFQGLXFSDTLLDF 182
 DB 120 YNNEILKQVDN-----KLLIYSIPALGTVALTTPQFAB--LFENPKIIGIXYNADFVLL 172
 QY 183 GQCVDQNRQQQAFLEFGVDEQLLSALVMGATGAVGS 218
 DB 173 ERV--RNAFPDKLILSGFDEMLPALALNVDCIGS 206
 RESULT 10
 DAPA HALK1 STANDARD; PRT; 304 AA.
 ID DAPA HALK1
 AC QHKS19;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDPS).
 GN DAPA OR VNG0444G.
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20504483; PubMed=11016950;
 RX NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Laaky S.R., Baliga N.S., Thorsson V., Sirogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough J.W., Dale H.,
 RA Madocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlshocher M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., Jassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -!- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =
 dihydrodipicolinate + 2 H(2O).
 CC -!- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
 semialdehyde; first step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the DHDPS family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: ALG95262; CAJ65647.1; -
 CC HAMAP: MF 01237; -; 1.
 CC InterPro: IPR002220; DHDPS.


```

DR EMBL: AE004999; AAG18989.1; -.
DR HAMAP; MF_00418; -. 1.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00731; DHDPS; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDPS; 1.
DR PROSITE; PS00665; DHDPS_1; FALSE_NEG.
DR PROSITE; PS00666; DHDPS_2; FALSE_NEG.
KW Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;
FT ACT_SITE 175 175 BY SIMILARITY.
KW Complete proteome.
FT ACT_SITE 175 175
SQ SEQUENCE 304 AA; 31397 MW; B1AP8826635116FC CRC64;

Query Match
Best Local Similarity 20.8%; Score 241; DB 1; Length 304;
Matches 72; Conservative 33; Mismatches 95; Indels 14; Gaps 7;

QY 8 LQGLVAATITPMTENGESVIGQYVYLVKEQGVKNIFVNGTIGEGLSLSVSERRQ 67
D5 16 LHGVPEVVTAPDAESLDADITADHA-RMWVDAGVHGVPPLGTNGEFPLLTPSERDRV- 73
QY 58 EHWTKGKXLD---QVIHVGALSLSKESCELAQHAETGADGIAVIAFFFLKPTWKDIL 124
D5 74 ---VTAVDEVGEVPIAGVAPSTRQ-VAAHRAHAASVAGDGVVVTTFYY-PLDGTAA 129
QY 125 INFLEKVEAAAPALPPYYHIALTGKVRABELLDGILDKIPTFOGLKFSDDTLDFGQ 184
D5 130 VEHYRVRVAANDC-PYVTVHIFSKTNEJSLFVL--AAAEIDTLAGVXKSDKVPWLQ 186
QY 185 CYDNRQQQFAFLFGVDEQLLSALVMGATGAVGS 213
D5 187 AVDAH--PELTFLAGSDSLAPGLDVGACGLVSA 213

RESULT 11
NANA_HAEIN STANDARD; PRT; 293 AA.
AC P44539;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE N-acetylneuraminatase (EC 4.1.3.3) (N-acetylneuraminic acid
DE aldolase) (N-acetylneuraminatase pyruvate-lyase) (Sialic acid lyase)
DE (Sialate lyase) (Sialic acid aldolase).
GN NANA OR H10142.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne C.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.W., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geckhagen N.S.M.,
RA Gnehm C.J., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512 (1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF NATIVE PROTEIN AND COMPLEXES
RP WITH SIALIC ACID ALDITOL; 4-DEOXY-SIALIC ACID AND 4-OXO-SIALIC ACID.
RX MEDLINE=20487594; PubMed=11031117;
RA Barbosa J.A.R.G., Smith B.J., DeGori R., Ooi H.C., Marcuccio S.M.,
RA Campi E.M., Jackson W.R., Brosmer R., Sommer M., Lawrence M.C.;
RT "Active site modulation in the N-acetylneuraminatase lyase subfamily as
RT revealed by the structure of the inhibitor-complexed Haemophilus

```

```

RT influenzae enzyme.";
RL J. Mol. Biol. 303:405-421 (2000).
CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
CC (sialic acid) to form pyruvate and N-acetylmannosamine via a
CC Schiff base intermediate.
CC -!- CATALYTIC ACTIVITY: N-acetylneuraminatase = N-acetyl-D-mannosamine +
CC pyruvate.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the DHDPS family. Nana subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.
CC
DR EMBL; U132700; AAC21814.1; -.
DR PIR; G64050; G64050.
DR PDB; 1F52; 20-DEC-00.
DR PDB; 1F6K; 20-DEC-00.
DR PDB; 1F6P; 22-NOV-00.
DR PDB; 1F73; 06-DEC-00.
DR PDB; 1F74; 22-NOV-00.
DR PDB; 1F7B; 22-NOV-00.
DR TIGR; H10142; -.
DR HAMAP; MF_01237; -. 1.
DR InterPro; IPR002220; DHDPS.
DR InterPro; IPR005264; Nana.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDPS; 1.
DR TIGRPFAM; TIGR00693; nana; 1.
DR PROSITE; PS00665; DHDPS_1; 1.
DR PROSITE; PS00666; DHDPS_2; 1.
KW Carbohydrate metabolism; Lyase; Schiff base; 3D-structure;
KW Complete proteome.
FT ACT_SITE 136 136 INVOLVED IN PROTON TRANSFER DURING
FT ACT_SITE 164 164 CLEAVAGE.
FT ACT_SITE 293 AA; 32564 MW; ACA755D0A5D9D33 CRC64;
SQ SEQUENCE 293 AA; 32564 MW; ACA755D0A5D9D33 CRC64;

Query Match
Best Local Similarity 20.8%; Score 240.5; DB 1; Length 293;
Matches 66; Conservative 44; Mismatches 81; Indels 37; Gaps 6;

QY 6 KKQLGLVAATITPMTENGESVIGQYVYLVKEQGVKNIFVNGTIGEGLSLSVSERRQ 65
D5 2 RDLKGIESALLVSFNEEDGTINEKGLRQIIRHNDIKMKVDGLYGVSGSTGENFMLSTEEK 61
QY 66 VAEWWTGKXLDQVIHVGALSLSKESCELAQHAETGADGIAVIAFFFLKPTWKDIL 125
D5 62 IFRIADKADQI-ALLAQGVSNLKEAVELGKYATELGYDCLSAVTPFYK----- 112
QY 126 NLFKEVAAAPALPVY-----YHPIALTGVKIRAEELDGILDKIPTFOGL 172
D5 113 -----FSFPEIKHYITITIAETGNMIVISIPFLTGVNMGIEQF--GELYKNPKVLG 163
QY 173 KFSDDT--ILDFGQCVQDNRRQQFAFLFGVDEQLLSALVMGATGAVGS 218
D5 164 KTAGDFVILLERLKKAYNH---LIWAGFDEMPLPAASLGVGDGAIGS 207

RESULT 12
NANA_RHIME STANDARD; PRT; 299 AA.
AC Q92WF0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE N-acetylneuraminatase lyase (EC 4.1.3.3) (N-acetylneuraminic acid

```



```

DE aldolase) (N-acetylneuraminase, pyruvate-lyase) (Sialic acid lyase)
DE (Sialate lyase) (Sialic acid aldolase).
GN NANA OR RBO289 OR SMB20299.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OS Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=1021;
RC MEDLINE=21396508; PubMed=1481431;
RX Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vortelschler F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-
RL fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -!- FUNCTION: Catalyzes the cleavage of N-acetylneuraminic acid
CC (sialic acid) to form pyruvate and N-acetylmannosamine via a
CC Schiff base intermediate (By similarity).
CC -!- CATALYTIC ACTIVITY: N-acetylneuraminate +
CC pyruvate.
CC -!- PATHWAY: N-acetylneuraminic acid (sialic acid) utilization.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHDPS family. Nana subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL603642; CAC48689.1; -
DR PIR; A95878; A95878.
DR HAMAP; MF 01237; -; 1.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR PRODOM; PD001859; DHDPS; 1.
DR PROSITE; PS00665; DHDPS_1; FALSE NEG.
DR PROSITE; PS00666; DHDPS_2; FALSE NEG.
KW Carbohydrate metabolism; Lyase; Schiff base; Plasmid;
KW Complete proteome.
FT ACT_SITE 134 134 INVOLVED IN PROTON TRANSFER DURING
FT ACT_SITE 161 161 CLEAVAGE (BY SIMILARITY).
FT ACT_SITE 161 161 SCHIFF-BASE WITH N-ACETYLNEURAMINATE
FT (BY SIMILARITY).
SQ SEQUENCE 299 AA; 32380 MW; 0FA610AD53886851 CRC64;
Query Match 20.5%; Score 237.5; DB 1; Length 299;
Best Local Similarity 29.2%; Pred. No. 7e-12;
Matches 62; Conservative 49; Mismatches 92; Indels 9; Gaps 7;
Qy 7 KLOGLVAATTPMTENGESVIGQYVDYLVKEQGVKNIFVNGTTGEGLSLSVSRQW 66
Db 2 KLEGIYSALLTFPESEDSIDRQALCALVDEQVR-LGIDGVYVGGSGEAMQLQSLDERADY 60
Qy 67 AEEWVTGKDKLDQVIVHVGALSLSKESQELAQHAAEIGADGIAVIAPFLKPDWTKDLIN 126
Db 61 LSDVAARASGRL-TLIAEVGTRIARDLRSLSHAQKSGYQALSIPIPPYDFSRPEVNAH 119
Qy 127 FLKEVAAAAPALPFYYHIPALTGKIRAEELLDGILDKIPTFQGLKFSDDTLDFGQCV 186
Db 120 Y-RELADVS-ALPIVYNFARTS-GFTLPPELVE--LLSHNPI-GIKHTSSDMQLERI- 173
Qy 187 DONFQQQAFALFGVDEQLLSALVMGATGAVS 218
Db 174 -RHAVPDAIVNGYDEMCLAGFAMGAQGAIGT 204

```

```

RESULT 13
DAPA METKA STANDARD; PRT; 300 AA.
ID DAPA METKA
AC Q8TUZ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydropicolinate synthase (EC 4.2.1.52) (DHDPS).
GN DAPA OR MKI607.
OS Methanopyrus kandleri.
OS Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyrazaceae;
OX Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RC MEDLINE=21927647; PubMed=11930014;
RX Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Koonin E.V., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -!- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =
CC dihydropicolinate + 2 H(2)O.
CC -!- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC semialdehyde; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the DHDPS family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE010450; AAM02823.1; -
DR HAMAP; MF 00418; -; 1.
DR InterPro; IPR005263; Dapa.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR PRODOM; PD001859; DHDPS; 1.
DR TIGRFAMs; TIGR00674; dapa; 1.
DR PROSITE; PS00665; DHDPS_1; 1.
DR PROSITE; PS00666; DHDPS_2; 1.
KW Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;
KW Complete proteome.
FT ACT_SITE 166 166 BY SIMILARITY.
SQ SEQUENCE 300 AA; 32345 MW; 42EFA59FE0D8883 CRC64;
Query Match 20.2%; Score 234.5; DB 1; Length 300;
Best Local Similarity 32.2%; Pred. No. 1.2e-11;
Matches 68; Conservative 42; Mismatches 94; Indels 7; Gaps 6;
Qy 7 KLOGLVAATTPMTENGESVIGQYVDYLVKEQGVKNIFVNGTTGEGLSLSVSRQW 65
Db 4 QIEGVIPALITPFTDGLKGINEEGLRENVSRL-LAAGVGVVPAGTTGESSTLSHAERR 62
Qy 66 VAESWVTGKDKLDQVIVHVGALSLSKESQELAQHAAEIGADGIAVIAPFLKPDWTKDLIN 125
Db 63 VIETWDEVNGKV-FVIAGAGSNSTREALELSTYAEDVGDAILLSWPYVYKPKQEGEFL 121
Qy 126 NFLKEVAAAAPALPFYYHIPALTGKIRAEELLDGILDKIPTFQGLKFSDDTLDFGQCV 185
Db 122 HFSK--TAEAVECEIILYNFSPRTGCALEPTAAK-LAEYSYHVGKERSGD-LDVQR 177
Qy 186 VDQNRQQQAFALFGVDEQLLSALVMGATGAV 216

```

178 FIETPPDFILLSSGVDELTL2LAUWGGVVI 208

Db

RESULT 14

DAPA METTH

ID DAPA METTH STANDARD; PRT; 283 AA.

AC 026892;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDP5).

GN DAPA OR MTH801.

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanothermobacter.

OC NCBI_TaxID=187420;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Delta H;

RA MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,

RA Aldredge T., Bashirzadeh R., Slakely D., Cook R., Gilbert K.,

RA Harrison D., Hoang L., Keagle P., Lum M., Pothier B., Qiu D.,

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,

RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,

RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,

RA Daniels C.-J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,

RA "Complete genome sequence of Methanobacterium thermoautotrophicum

RT deltaH: functional analysis and comparative genomics.";

RL J. Bacteriol. 179:7135-7155(1997).

CC -1- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =

CC dihydrodipicolinate + 2 H(2)O

CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate

CC semialdehyde; first step.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- SIMILARITY: Belongs to the DHDP5 family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC

EMBL; A69207; A69207.

DR PIR; A69207; A69207.

DR HSSP; PC5640; 1DHP.

DR HAMAP; MF_00418; -; 1.

DR InterPro; IPR005263; DAPA.

DR Pfam; PF00701; DHDP5; 1.

DR PRINTS; PR00146; DHPICSNTHASE.

DR ProDom; PD001859; DHDP5; 1.

DR TIGRfam; TIGR00674; dapa; 1.

DR PROSITE; PS00665; DHDP5_1; 1.

DR PROSITE; PS00666; DHDP5_2; 1.

DR Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;

KW Complete proteome.

KW ACT SITE 162 162

FT ACT SITE 162 162

SQ SEQUENCE 283 AA; 30362 MW; 8F71C2E574D83B64 CRC64;

Query Match 20.0%; Score 231.5; DB 1; Length 283;

Best Local Similarity 27.3%; Pred. No. 2e-11;

Matches 62; Conservative 52; Mismatches 90; Indels 23; Gaps 7;

QY 7 KLQGLVAATITPMTENGSEINFSVIGQVYLVKEQGVNIFVNGITTGEGLSVSRERQV 66

Db 2 KEGTVAWVTPEDDVVDREAGLRNINXLI-ENGVDGLLVAGITGESATITHEEQRM 60

QY 67 AEWVTGKDKLDQWIIHVGALSLSKESQBLAQAHAIGADGIANIAPFLKPKWKDILLIN 126

Db 61 IDILVDEVNGRV-RTVAGAGSNSREAMGLVEYAEADAGADAALVITPYNKPQPHGLIEH 119

127 F--LKEVAAAAPALPFYVYHIPALTYGKIRAEEL-----LDGIL---DKIETFOGLKFSF 176

Db YMLLEFAA-----DPLIILYVPSRTGTVDVTAELAKLGLIGIKASPDLDKVSMLR 175

177 TDLIDFGQCVDPNRQQQFAFLFGVDEQLLSALVMGATGAVGSFVSRD 223

Db SRLMDLG-----LDDFTVLSGNDLTLPLMISMGAGVISVVAUVD 215

RESULT 15

DAPA METMA

ID DAPA METMA STANDARD; PRT; 291 AA.

AC Q8PX17;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Dihydrodipicolinate synthase (EC 4.2.1.52) (DHDP5).

GN DAPA OR MM1201.

OS Methanosarcina mazei (Methanosarcina frisia).

OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;

OC Methanosarcinaceae; Methanosarcina.

OC NCBI_TaxID=2209;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Goei / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;

RA MEDLINE=22120827; PubMed=12125824;

RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,

RA Martinez-Arias R., Henne A., Waezler A., Baumer S., Jacobi C.,

RA Brueggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,

RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,

RA Fritz H.-J., Gottschalk G.;

RA "The genome of Methanosarcina mazei: evidence for lateral gene

RT transfer between Bacteria and Archaea.";

RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).

CC -1- CATALYTIC ACTIVITY: L-aspartate 4-semialdehyde + pyruvate =

CC dihydrodipicolinate + 2 H(2)O

CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate

CC semialdehyde; first step.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- SIMILARITY: Belongs to the DHDP5 family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC

EMBL; A6913350; AAM30897.1; -.

DR HAMAP; MF_00418; -; 1.

DR InterPro; IPR005263; DAPA.

DR InterPro; IPR002220; DHDP5.

DR Pfam; PF00701; DHDP5; 1.

DR PRINTS; PR00146; DHPICSNTHASE.

DR ProDom; PD001859; DHDP5; 1.

DR TIGRfam; TIGR00674; dapa; 1.

DR PROSITE; PS00665; DHDP5_1; 1.

DR PROSITE; PS00666; DHDP5_2; 1.

DR Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;

KW Complete proteome.

KW ACT SITE 159 159

FT ACT SITE 159 159

SQ SEQUENCE 291 AA; 30704 MW; 45BFB0BD0C17FFA26 CRC64;

Query Match 19.5%; Score 226; DB 1; Length 291;

Best Local Similarity 28.8%; Pred. No. 5.7e-11;

Matches 60; Conservative 44; Mismatches 96; Indels 8; Gaps 4;

QY 9 QGLVAATITPMTENGSEINFSVIGQVYLVKEQGVNIFVNGITTGEGLSVSRERQV 68

Db 3 EGMPALITPFDKDRIDREGLQRIAF-VEEGVSGIVPCGTGTGESATLSAAHEVID 61

Qy	69	EWWTKGKXLDQVLIHWGALSLSKESQELAQHAEIGADSI	AVIAPFFLKPTKDILINFL	128
Db	62	IAVECSK--VPVIAGTGSNNTGEALQFTKHAADAGVDG	VLLISPYNKNPNPAGLLAHFK	118
Qy	129	KEVAAAAPALPFYVYHIPALTGVKIRAEELLOGILD	KIPTFOGLKFSDTDLIDFGQCVDQ	188
Db	119	X--IAEAVDIEWILYNVESRTQDMPVDVIVE--LAK	VENIVGIKEASGNAKVSQILEN	174
Qy	189	NRQQQFAPLFGVDEQLLSALVNGATGAV		216
Db	175	TMDDDFVVLSCEDGTLPLISNGGRGVI		202

Search completed: May 6, 2004, 09:04:58
 Job time : 7.04821 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:19 ; Search time 22.5128 Seconds
(without alignments)
3223.466 Million cell updates/sec

Title: US-09-930-440b-2
Perfect score: 1159
Sequence: 1 MAPPKKQLGLVAATITPMT.....GATGAVGFSVRSJLSTLSN 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1104.5	95.3	320	4 Q9BXD5	Q9bxds5 homo sapien
2	1032.5	89.1	319	6 Q9BEC7	Q9bec7 sus scrofa
3	998.5	86.2	320	11 Q9DCQ9	Q9dcq9 mus musculus
4	443	38.2	297	16 Q8ZCG7	Q8zcg7 yersinia pe
5	441	38.1	299	16 Q8DOX1	Q8dxx1 yersinia pe
6	430	37.1	298	16 Q9KR67	Q9kr67 vibrio chol
7	421.5	36.4	322	16 Q7UUE0	Q7uue0 rhodospirill
8	399	34.4	299	16 Q8D617	Q8d617 vibrio vuln
9	273	23.6	306	2 Q8RR42	Q8rr42 streptococc
10	256	22.1	305	16 Q97P53	Q97p53 streptococc
11	252.5	21.8	305	16 Q97Q96	Q97q96 streptococc
12	251.5	21.7	294	17 Q9UZ94	Q9uz94 pyrococcus
13	248.5	21.4	318	5 Q27818	Q27818 trichomonas
14	247	21.3	305	16 Q8E7V7	Q8e7v7 streptococc
15	247	21.3	305	16 Q8E2E9	Q8e2e9 streptococc
16	243	21.0	226	2 Q32375	Q32375 clostridium

17	232	20.0	289	16 Q83CA6	Q83ca6 coxiella bu
18	232	20.0	302	17 Q97UF0	Q97uf0 sulfolobus
19	231.5	20.0	289	17 Q97WF2	Q97wf2 sulfolobus
20	230.5	19.9	287	17 Q58577	Q58577 pyrococcus
21	222	19.2	298	16 Q81CD9	Q81cd9 bacillus ce
22	220	19.0	298	17 Q8U319	Q8u319 pyrococcus
23	219	18.9	298	16 Q81PH3	Q81ph3 bacillus an
24	218.5	18.9	292	16 Q81WN7	Q81wv7 bacillus an
25	218.5	18.9	232	16 Q819Z9	Q819z9 bacillus ce
26	218	18.8	304	16 Q8X8P0	Q8x8p0 streptococc
27	215	18.6	304	16 Q9A1I5	Q9a1i5 streptococc
28	215	18.6	304	16 Q8P2P3	Q8p2p3 streptococc
29	214	18.5	292	10 Q8H725	Q8h725 phytoptchor
30	211.5	18.2	301	16 Q89Q28	Q89q28 bradyrhizob
31	211	18.2	285	17 Q96YD3	Q96yd3 sulfolobus
32	211	18.2	302	16 Q7UA33	Q7ua33 synectococc
33	211	18.2	319	16 Q7UMJ1	Q7umj1 rhodospirill
34	210.5	18.2	302	16 Q7V590	Q7v590 prochloroco
35	209.5	18.1	295	16 Q8ANH2	Q8anh2 pseudomonas
36	209	18.0	308	16 Q8EMH5	Q8emh5 oceanobacil
37	208	17.9	302	17 Q97UI2	Q97ui2 sulfolobus
38	207.5	17.9	290	2 Q93RJ8	Q93rj8 methyllobaci
39	205.5	17.7	292	17 Q9HJ17	Q9hj17 thermoplasma
40	204.5	17.6	295	16 Q8FJW9	Q8fjw9 escherichia
41	202	17.4	297	16 Q7VM87	Q7vm87 haemophilus
42	197.5	17.0	292	16 Q82SD7	Q82sd7 nitrosomona
43	196.5	17.0	292	16 Q7VXZ8	Q7vxz8 bordetella
44	196.5	17.0	301	16 Q7W884	Q7w884 bordetella
45	195.5	16.9	294	2 Q7X387	Q7x387 escherichia

ALIGNMENTS

RESULT 1

ID	Q9BXD5	PRELIMINARY;	PRT;	320 AA.
AC	Q9BXD5	Q9BXD5		
DT	01-JUN-2001	(TREMREL. 17, Created)		
DT	01-JUN-2001	(TREMREL. 17, Last sequence update)		
DT	01-OCT-2002	(TREMREL. 22, Last annotation update)		
DE	Clorf13			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=2121827; PubMed=11318611;			
RA	Sood R., Bonner T.I., Malakowska I., Stephan D.A., Robbins C.M.,			
RA	Connors T.D., Morgenbesser S.D., Su K., Parague M.U., Pinkett H.,			
RA	Graham C., Baxevaris A.D., Klinger K.W., Landes G.M., Trent J.M.,			
RA	Carpten J.D.;			
RT	*Cloning and characterization of 13 novel transcripts and the human			
RT	RG88 gene from the lq25 region encompassing the hereditary prostate			
RT	cancer (hpc1) locus";			
RL	Genomics 73:211-222(2001).			
DR	EMBL; AF338436; AAK25795.1; -			
DR	Genew; HGNC:16781; Clorf13.			
DR	InterPro; IPR002220; DHDPS.			
DR	Pfam; PF00701; DHDPS; 1.			
DR	PRINTS; PR00146; DHDPSNTBASE.			
DR	ProDom; PD001859; DHDPS; 1.			
SQ	SEQUENCE 320 AA; 35162 MW; FC0EBA9B05FE9E62 CRC64;			

Query Match	95.3%;	Score 1104.5;	DB 4;	Length 320;
Best Local Similarity	91.0%;	Pred. No. 4.5e-86;		
Matches 222;	Conservative	2;	Mismatches 5;	Indels 15;
				Gaps 1;
QY	1	MAPPKKQLGLVAATITPMTENGEINFSVIGQVVDYLVKEQGVKNIFVNGTTGEGLSLV	60	
DB	1	MAPPKKQLGLVAATITPMTENGEINFSVIGQVVDYLVKEQGVKNIFVNGTTGEGLSLV	60	

Thu May 13 11:53:01 2004

```
QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
QY 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLISN 229
DB 241 SALN 244

RESULT 2
Q9BEC7 PRELIMINARY; PRT; 319 AA.
AC Q9BEC7;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 01-JUN-2001 (TRENBLrel. 24, Last annotation update)
DE Acylneuraminase lyase.
CN NPL
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Bruse P., Traving C., Schauer R.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC271330; CAC27797.1; -
DR GO; GO:0016229; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDP5.
DR Pfam; PF00731; DHDP5; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDP5; 1.
KW Lyase.
SQ SEQUENCE 319 AA; 35062 MW; C89894E9F68674A7 CRC64;

Query Match
Best Local Similarity 84.1%; Score 1032.5; DB 6; Length 319;
Matches 206; Conservative 13; Mismatches 11; Indels 15; Gaps 1;

QY 1 MAFPKKLGQGLVAATITPMTENGINFVSIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
DB 1 MAFPKKLGQGLVAATITPMTENGINFVSIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
QY 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLISN 230
DB 241 SALN 245

RESULT 3
Q9BEC9 PRELIMINARY; PRT; 320 AA.
ID Q9BEC9
AC Q9BEC9;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 01-JUN-2001 (TRENBLrel. 24, Last annotation update)
DE Acylneuraminase lyase.
CN NPL
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Bruse P., Traving C., Schauer R.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC271330; CAC27797.1; -
DR GO; GO:0016229; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDP5.
DR Pfam; PF00731; DHDP5; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDP5; 1.
KW Lyase.
SQ SEQUENCE 319 AA; 35062 MW; C89894E9F68674A7 CRC64;

Query Match
Best Local Similarity 84.1%; Score 1032.5; DB 6; Length 319;
Matches 206; Conservative 13; Mismatches 11; Indels 15; Gaps 1;

QY 1 MAFPKKLGQGLVAATITPMTENGINFVSIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
DB 1 MAFPKKLGQGLVAATITPMTENGINFVSIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
QY 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLISN 230
DB 241 SALN 245
```

```
01-JUN-2001 (TRENBLrel. 17, Created)
01-JUN-2001 (TRENBLrel. 17, Last sequence update)
01-MAR-2003 (TRENBLrel. 23, Last annotation update)
CG-0033B02Rik protein (RIKEN cDNA 0610033B02 Gene)
DE (N-acetylneuraminase pyruvate lyase).
CN NPL OR 0610033B02Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
MEDLINE=2108566; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino Y., Itch M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojibori T., Soto H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Monbaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RL Nature 420:563-573(2002).
RA "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK002734; BAB22314.1; -
DR EMBL; BC022734; AAB22734.1; -
DR EMBL; AK088859; BAC40618.1; -
DR MGD; MGI:1921341; NpL.
DR InterPro; IPR002220; DHDP5.
DR Pfam; PF00701; DHDP5; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDP5; 1.
SQ SEQUENCE 320 AA; 35130 MW; 9426AD7CC8438468 CRC64;

Query Match
Best Local Similarity 86.2%; Score 998.5; DB 11; Length 320;
Matches 198; Conservative 17; Mismatches 14; Indels 15; Gaps 1;

QY 1 MAFPKKLGQGLVAATITPMTENGINFVSIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
DB 1 MAFPKKLGQGLVAATITPMTENGINFVSIGQVYDVLVKEQGVKNIPVNGTTGEGLSLV 60
QY 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
DB 61 SERQVAEEWVTGKDKLDQVLIHVGLSLKESQELAQHAAEIGDGLAVIAPFLKPWT 120
QY 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
DB 121 KDILINFLKEVAAAPALPFYVYHHPALTGKIRAEILLDGLDKIPTFQGLKESDIDL 180
QY 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
DB 181 DFGQCVQNRQQQFAFLFGVDEQILLSALVNGATGAVGS-----FVSRDLS 225
QY 226 TLISN 230
DB 241 SALN 245
```

Thu May 13 11:53:01 2004

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative acetylneuraminase lyase.
 GN Y1458.
 OS Versinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 EX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.,
 RT "Genome sequence of Yersinia pestis KIM."
 RL J. Bacteriol. 184:4501-4611(2002).
 DR EMBL; AB013749; AM85029.1;
 DR GO; GO:0016829; F-lyase activity; IEA.
 DR InterPro; IPR002220; DHDPS.
 DR Pfam; PF00701; DHDPS_1.
 DR PRINTS; PR00146; DHDPS_1.
 DR PRODOM; PD001859; DHDPS_1.
 DR PROSITE; PS00665; DHDPS_1; 1.
 DR Lyase.
 SQ SEQUENCE 299 AA; 32179 MW; 846EBF6165791575 CRC64;
 KW
 Query Match 38.1%; Score 441; DB 16; Length 299;
 Best Local Similarity 43.7%; Pred. No. 1.3e-29;
 Matches 93; Conservative 40; Mismatches 76; Indels 4; Gaps 4;
 QY 6 KKLQGLVAATITPTMTENGESVIGQYDYLKVEQGVNIFVNGTGTGEGLSLSVSRQ 65
 DB 4 KKLTLGLIAAPHTPFDEQGVNYPVIDQIAHLIND-GVKGYYVCGTGTGEGHCSVDERKK 62
 QY 66 VAEWTKGDKLDQVLIHVGALSKESELAQAAEIGADGIAVIAPFFLKPTKIDILI 125
 DB 63 IAEWYNAAGKGL-SITLHTGALSIKDAVLSRHAETLDFATSAIGCFKFGNLDLI 121
 QY 126 NFLKEVAAAAPALPFYHYHYPALTVKIRAEELLDGILDKIPTFGSKFSDTLLDFGQC 185
 DB 122 AYCQAIATAAPSKGFYYH-SGMGVNLDMEQFLIKAESKIPNLGKFNADLYEFQRC 180
 QY 186 VQNRQQQFAFLFGVDEQLLSALVMGATGAVS 218
 DB 181 L-RVSGGKFDIPFGVDEHLPGGLAVGATGAVS 212
 RESULT 5
 Q9KR67 PRELIMINARY; PRT; 298 AA.
 ID Q9KR67
 AC Q9KR67
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE N-acetylneuraminase lyase, putative.
 GN VCI776.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=E1 Tor N16961 / Serotype O1;
 EX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unwayan L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Baes S., Qin H., Diago I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

QY 181 DFGQCVQDNRQQQFAFLFGVDEQLLSALVMGATGAVS-----FVSRDL 225
 DB 181 DFGQCVQDNRQQQFAFLFGVDEQLLSALVMGATGAVS-----FVSRDL 240
 QY 226 TLLS 229
 DB 241 SALS 244
 RESULT 4
 Q8ZG37 PRELIMINARY; PRT; 297 AA.
 ID Q8ZG37
 AC Q8ZG37
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Probable N-acetylneuraminase lyase (EC 4.1.3.3).
 GN YP03024.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 EX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 DR EMBL; AJ14155; CAC92266.1; -.
 DR PIR; AG0367; AG0367.
 DR GO; GO:0016829; F-lyase activity; IEA.
 DR GO; GO:0008747; F-N-acetylneuraminase lyase activity; IEA.
 DR InterPro; IPR002220; DHDPS.
 DR Pfam; PF00701; DHDPS_1.
 DR PRINTS; PR00146; DHDPS_1.
 DR PRODOM; PD001859; DHDPS_1.
 DR PROSITE; PS00665; DHDPS_1; 1.
 DR Lyase; Complete proteome.
 KW SEQUENCE 297 AA; 31950 MW; 57C1D709D25514C5 CRC64;
 Query Match 38.2%; Score 443; DB 16; Length 297;
 Best Local Similarity 43.7%; Pred. No. 1.3e-29;
 Matches 93; Conservative 41; Mismatches 75; Indels 4; Gaps 4;
 QY 6 KKLQGLVAATITPTMTENGESVIGQYDYLKVEQGVNIFVNGTGTGEGLSLSVSRQ 65
 DB 2 KKLTLGLIAAPHTPFDEQGVNYPVIDQIAHLIND-GVKGYYVCGTGTGEGHCSVDERKK 60
 QY 66 VAEWTKGDKLDQVLIHVGALSKESELAQAAEIGADGIAVIAPFFLKPTKIDILI 125
 DB 61 IAEWYNAAGKGL-SITLHTGALSIKDAVLSRHAETLDFATSAIGCFKFGNLDLI 119
 QY 126 NFLKEVAAAAPALPFYHYHYPALTVKIRAEELLDGILDKIPTFGSKFSDTLLDFGQC 185
 DB 120 AYCQAIATAAPSKGFYYH-SGMGVNLDMEQFLIKAESKIPNLGKFNADLYEFQRC 178
 QY 186 VQNRQQQFAFLFGVDEQLLSALVMGATGAVS 218
 DB 179 L-RVSGGKFDIPFGVDEHLPGGLAVGATGAVS 210
 RESULT 5
 Q8D0X1 PRELIMINARY; PRT; 299 AA.
 ID Q8D0X1
 AC Q8D0X1

```
RA Salzberg S.B., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Praser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RJ Nature 406:477-483 (2000).
DR EMBL; AB004255; AAF94925.1; -.
DR FIR; B82157; B82157.
DR TIGR; VCI1776; -.
DR GO; GO:0016829; F-lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDPS; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 298 AA; 32221 MW; 4359567DC749BB6D CRC64;

Query Match 37.1%; Score 430; DB 16; Length 298;
Best Local Similarity 44.1%; Pred. No. 1.7e-28;
Matches 94; Conservative 39; Mismatches 76; Indels 4; Gaps 4;

QY 6 KKQGLVAAATTPMTENGESVIGQYDYLKQGVKNIFVNGTTGEGLSLSVSERRQ 65
Db 2 KKLGLIAAPHTFTKKNKFAIDQIAELL-EGQVGAYVCGTTGEGIHCSVERKA 60

QY 66 VAEWWTGKDKLDQVTHVHGSLSKESQELAQHAAEIGADGIAVIAPFFKPTWKDILI 125
Db 61 IASRWKAVDQKLD-VLHTGALSIVDTLHTHAETLDIFATSAIGPCFFKPSVVDLV 119

QY 126 NELKEVAAAAPALPFYYHHPALTGVKIRAEELDGLDKIPTFGSLKFSDDLDFGQC 185
Db 120 EYCAQVAAAAPSKGFYYH-SGSGVNLDEOFLKGEQRIPLNYGAKFNADLYEQRC 178

QY 186 VDQNRQOQFAFLFGVDQLLSALVMGATGAVGS 218
Db 179 V-RVSNKRFIDIPFGVDEFLPAGLAVGATGAVGS 210

RESULT 7
Q7JU60 PRELIMINARY; PRT; 322 AA.
AC Q7JU60;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25; Last sequence update)
DE Probable N-acetylneuraminatase lyase (EC 4.1.3.3).
GN NANA OR RH3352.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=20735913; PubMed=12835416;
RA Gloeckner F.O., Kube W., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RC "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294138; CND73140.1; -.
KW Lyase; Complete proteome.
SQ SEQUENCE 322 AA; 34857 MW; 02BAC02AB21AA68C CRC64;

Query Match 36.4%; Score 421.5; DB 16; Length 322;
Best Local Similarity 43.7%; Pred. No. 9.8e-28;
Matches 93; Conservative 41; Mismatches 76; Indels 3; Gaps 3;

QY 6 KKQGLVAAATTPMTENGESVIGQYDYLKQGVKNIFVNGTTGEGLSLSVSERRQ 65
Db 15 RKLGLIAATYPMKACGDLNDLVPMAYEKLLD-GVSGLVCGSTGEGMSLTTRERQL 73

QY 66 VAEWWTGKDKLDQVTHVHGSLSKESQELAQHAAEIGADGIAVIAPFFKPTWKDILI 125
```

```
Db 74 VASAFVNATAGRV-PVIVQGENSLAEQELAKHAQOIGASALSATCSYFKVASQSLT 132
QY 126 NFLKEVAAAAPALPFYYHHPALTGVKIRAEELDGLDKIPTFGSLKFSDDLDFGQC 185
Db 133 LCMQELAAAAAPETFFYYHHPALTGVKIRAEELDGLDKIPTFGSLKFSDDLDFGQC 192
QY 186 VDQNRQOQFAFLFGVDQLLSALVMGATGAVGS 218
Db 193 LELSL-NRKEDVWGCDEMILLGATATGARAAGS 224

RESULT 8
Q8D617 PRELIMINARY; PRT; 239 AA.
AC Q8D617;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23; Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24; Last annotation update)
DE Dihydrodipicolinate synthase/N-acetylneuraminatase lyase.
GN VV20730.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choe H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (JBC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEC16810; AAC07662.1; -.
DR GO; GO:0016829; F-lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PR00146; DHPICSNTHASE.
DR ProDom; PD001859; DHDPS; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 239 AA; 32199 MW; 69CEC84008AB211A CRC64;

Query Match 34.4%; Score 399; DB 16; Length 299;
Best Local Similarity 41.5%; Pred. No. 7.4e-26;
Matches 88; Conservative 40; Mismatches 80; Indels 4; Gaps 4;

QY 7 KQGLVAAATTPMTENGESVIGQYDYLKQGVKNIFVNGTTGEGLSLSVSERRQV 66
Db 4 KKLGLIAAPHTFTKKNKFAIDQIAELL-EGQVGAYVCGTTGEGIHCSVERKA 62

QY 67 AEEWWTGKDKLDQVTHVHGSLSKESQELAQHAAEIGADGIAVIAPFFKPTWKDILI 126
Db 63 AERWVKAADGKLD-IILHTGALSIVDTLHTHAETLDIFATSAIGPCFFKPSVVDLVN 121

QY 127 FLKEVAAAAPALPFYYHHPALTGVKIRAEELDGLDKIPTFGSLKFSDDLDFGQC 186
Db 122 YCAQTAFAAPSKGFYYH-SGSGVNLDEOFLKGEQRIPLNYGAKFNADLYEQRCAL 180

QY 187 DQNRQOQFAFLFGVDQLLSALVMGATGAVGS 218
Db 181 -RVSNKRFIDIPFGVDEFLPAGLAVGATGAVGS 211

RESULT 9
Q8RR42 PRELIMINARY; PRT; 305 AA.
AC Q8RR42;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21; Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25; Last annotation update)
DE Putative N-acetylneuraminatase lyase.
GN NANA.
OS Streptococcus suis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
```



```

OC Streptococcus.
ON NCBI_TaxID=1307;
RX [1];
RP SEQUENCE FROM N.A.
RC STRAIN=10581; PubMed=11889115;
RX MEDLINE=21386400; Sekizaki T.;
RT Takamatsu D., Osaki M., Sekizaki T.;
RT "Evidence for Lateral Transfer of the sulfolysin Gene Region of
RT Streptococcus suis.";
RL J. Bacteriol. 184:2050-2057(2002).
DR EMEL; AB071365; BAB85970.1; -
DR GO: GO:0016829; F:lyase activity; IEA.
DR InterPro: IPR002220; DHDPS.
DR InterPro: IPR000437; Prok_lipoprot S.
DR Pfam: PF00701; DHDPS; 1.
DR PRINTS: PR00146; DHDPS; 1.
DR PRODOM: PD001859; DHDPS; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR Lyase.
SQ SEQUENCE 306 AA; 33402 MW; 038A876FE750356D CRC64;

Query Match 23.6%; Score 273; DB 2; Length 306;
Best Local Similarity 33.6%; Pred. No. 4.4e-15;
Matches 72; Conservative 41; Mismatches 89; Indels 12; Gaps 6;

QY 7 KXQGLVAATITPMTENGINFVSIGQYDYLVKEQGVKNIFVNGTGTGEGLSLSVSERRQV 66
DB 6 KXHGVIIPAFYACVDRAGEISSESVKVLQVFT-DKGVQGLYVNGSGSGECIYQSVADRKQI 64

QY 67 AEENWTKGDKLDQVLIHVGLSLKESQELAQHAAEIGADGIAVIAPFFLKPTWKDIL 126
DB 65 LEAVMEVAKGKL-TIIHVACNNKDSVELARHSEELGVDAIAIPPIYFR-LPEYSVA 122

QY 127 FLKEVAAAAPALPFYVYHIPALTGVKIRAE---ELLDGILDKIPTFGQLKFSDTLLDFG 183
DB 123 YWNGISAAAPNTDFLIYNIPQLAGVSLTSLYKEML-----KNPRVGVKNSSMPVQDI 177

QY 184 QCVDMNRQQQFAFLFGVDQLLSALVNGATGAVG 217
DB 178 TFFVLGGEDYVVF-NGPDEQFLGRLMGARAGIG 210

RESULT 10
Q972F3 PRELIMINARY; PRT; 305 AA.
AC Q97PF3;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-acetylneuraminase lyase, putative.
GN SP1676.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1313;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfle E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMEL; AE007431; AAK75427.1; -
DR PIR; B95154; B95154.
DR TIGR; SP1329; -
DR GO: GO:0016829; F:lyase activity; IEA.
DR InterPro: IPR002220; DHDPS.
DR Pfam: PF00701; DHDPS; 1.
DR PRINTS: PR00146; DHDPS; 1.
DR PRODOM: PD001859; DHDPS; 1.
DR Lyase. Complete proteome.
KW SEQUENCE 305 AA; 33562 MW; 5F8A661DD5C86ED7 CRC64;

Query Match 21.8%; Score 252.5; DB 16; Length 305;
Best Local Similarity 31.1%; Pred. No. 2.5e-13;
Matches 73; Conservative 37; Mismatches 104; Indels 21; Gaps 6;

QY 7 KXQGLVAATITPMTENGINFVSIGQYDYLVKEQGVKNIFVNGTGTGEGLSLSVSERRQV 66

```

Thu May 13 11:53:01 2004

us-09-930-440b-2.rspt

```

Db      6 KYGVIPAFACYDNGEISQDRKSLVQYFI-DKGVKGIYVNGSGEGCIYVSVDREKQI 64
QY      67 AEWWTGKQKLDQVLIHVGALSKEQELAGADGAGIAPFFLKPTWKDILIN 126
Db      65 IZAVMEVAKGKI-TVINHACANTKDSIELAKHESVGVDAJAAIPIYFK-LPEYSIAA 122
QY      127 FLKEVAAAAPALPFYYHYHPIATGVKIRAEELLDGILKIPFQGLKFSDDTLDDPGQCV 186
Db      123 YNANSEAAASNTDFIYINPLAGVALTGS--LYATMRQNPRIYGVKNSMPVQDIDQXVF 180
QY      187 DQNRQOQAFELFGVDEQLLSALVWGATGAVS-----FVSRLST 226
Db      181 AAGGEDYVVF-NGEEDCYLGGRLMGABAGIGYGVMPDLFLKLESILQERDLD 234

RESULT 12
Q9UZ94 PRELIMINARY; PRT; 294 AA.
AC Q9UZ94;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Dihydrodipicolinate synthase (DAP).
GN DAPA OR PYRAB1600 OR PAB0832.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
CC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GBS / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EXBL; AJ248297; CAB50165.1; -.
DR PIR; H75033; H75033.
DR InterPro; IPR002220; DHPS.
DR Pfam; PF00701; DHPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR ProDom; PD001859; DHPS; 1.
DR Complete Proteome.
KW Complete Proteome.
SQ SEQUENCE 294 AA; 32913 MW; 7FA3D40BF9E9FAC2 CRC64;

Query Match 21.7%; Score 251.5; DB 17; Length 294;
Best Local Similarity 30.1%; Pred. No. 2.9e-13;
Matches 65; Conservative 49; Mismatches 81; Indels 21; Gaps 7;

QY 8 LQGLVAATITPMTENGEINFSVIGQYDYLKQGVKNIFVNTTGEGLSLSVSRROVA 67
Db 1 MEGVIVELVTFREDYSDIFALEWHIREL--EEKGVHGIIVNSTTGETSINTDERKLA 59
QY 68 EEW-VTKGDKLDQVLIHVGALSKEQELAGADGAGIAPFFLKPTWKDILIN 125
Db 60 EKGREITSG-----MYLVGTSTSTLEVLISRAEDIGADGIV-VAPYCKLKDEILK 114
QY 126 NPLKEVAAAAPALPFYYHYHPIATGVKIRAEELLDGILKIPFQGLKFSDDTLDD 181
Db 115 HF--SMVAERVDIPYIVVAIPSC-----ANPVPVDIIRKVSLEYNSIIGKVASVDSJTY 166
QY 182 FGQCYD-QNRQOQAFELFGVDEQLLSALVWGATGAV 216
Db 167 -QELIEVKEERKDFRVFTGLDQYFJLTLGGDGI 202

RESULT 13
Q27818 PRELIMINARY; PRT; 318 AA.
AC Q27818;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

```

```

DE N-acetylneuraminase lyase (EC 4.1.3.3).
GN NANA.
OS Trichomonas vaginalis.
OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae;
OC Trichomonadinae; Trichomonas.
OX NCBI_TaxID=5722;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=clinical isolate #202;
RX MEDLINE=97077447; PubMed=8920014;
RA Meysick K.C., Dimock K., Garber G.B.;
RT "Molecular characterization and expression of a N-acetylneuraminase
lyase gene from Trichomonas vaginalis.";
RL Mol. Biochem. Parasitol. 76:289-292(1996).
RL EMBL; U35878; AAB42182.1; -.
DR HSSP; P06995; LFDY.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016029; F:lyase activity; IEA.
DR GO; GO:0008747; F:N-acetylneuraminase lyase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR002220; DHPS.
DR InterPro; IPR005264; Nana.
DR InterPro; IPR005829; Sug_transporter.
DR Pfam; PF00701; DHPS; 1.
DR PRINTS; PR00146; DHPICNTHASE.
DR ProDom; PD001859; DHPS; 1.
DR TIGRfams; TIGR00683; nana; 1.
DR PROSITE; PS00665; DHPS 1; 1.
DR PROSITE; PS00666; DHPS 2; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Lyase.
SQ SEQUENCE 318 AA; 35083 MW; 25F0244C298D593D CRC64;

Query Match 21.4%; Score 248.5; DB 5; Length 318;
Best Local Similarity 30.0%; Pred. No. 5.8e-13;
Matches 64; Conservative 47; Mismatches 95; Indels 7; Gaps 5;

QY 5 KKLQGLVAATITPMTENGEINFSVIGQYDYLKQGVKNIFVNTTGEGLSLSVSRRO 65
Db 27 KSLKGLFSALLVSPNEDGTINEKGLREIVRYNIDKIDGLYGVGSGTENFELSTEKKQ 86
QY 66 VAEWWTGKDKLDQVLIHVGALSKEQELAGADGAGIAPFFLKPTWKDILIN 125
Db 87 IFRIADKADQV-ALLAQVGSINIHESIEIGKVATELGYNCLSAVTFYFK-TFFPEIK 144
QY 126 NPLKEVAAAAPALPFYYHYHPIATGVKIRAEELLDGILKIPFQGLKFSDDTLDD 185
Db 145 NYITVIVAT-GNNMIVYISIPALTGVSWTADQF--GELFENPKIIGVFTAGDFYLLSRV 201
QY 186 VQNRQOQAFELFGVDEQLLSALVWGATGAVS 218
Db 202 --KRAYPDHLIWAQFDEMMLPACSGIDGAGS 232

RESULT 14
Q8E7V7 PRELIMINARY; PRT; 305 AA.
AC Q8E7V7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN GBS0038.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22425208; PubMed=12354221;
RA Glaser P., Rusznick C., Buchrieser C., Chevalier F., Frangeul L.,

```

RA Msadek T., Zouine M., Couve E., Lalicai L., Poyart C., Trieu-Cuot P.,
RT Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766843; CAD45683.1; -;
DR Sagalists; gbs0038; -;
DR InterPro; IPR002220; DHDPS.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PRO0146; DHPICSNTHASE.
DR ProDom; PD001859; LHDPS; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 305 AA; 33495 MW; 89E06287CAC18C15 CRC64;

Query Match 21.3%; Score 247; DB 16; Length 305;
Best Local Similarity 32.1%; Pred. No. 7.4e-13;
Matches 68; Conservative 40; Mismatches 98; Indels 6; Gaps 6;

QY 6 XKQGLVAAATITPMTENGSEINFSVIGQYVYLVKEQGVKNIFVNGTGTGEGLSLSVSERRQ 65
DB 5 QKVGIIIPAFYACYDDKGDICPERVKALTYFT-DKGVQGLYVNGSSGECIYQSVADKRL 63
QY 66 VAEWVTKGKDKLDQVTHVHGALSLSKESQELACHAAEIGADGIVIAPIPFELKPTWTKDILL 125
DB 64 VLENVMSVAKGKL-TVIAHVACNNTKDSVELAMHAEIGVDAIAALPTIYFR-LPEYALIA 121
QY 126 NFLKEVAAAPALPFYVYHIPALTVGKIRAEELDGLDKIPTFGQKFSDTDLDFGQC 185
DB 122 DYNTISQAAPQDTFIYINIPQAGVAL-TSDLYRKMLCN-PQVIGVKNSSMPVQDIQNF 179
QY 186 VQNRQQQFAFLFGVDEQLLSALVMGATGAVG 217
DB 180 VAIGGENHIVF-NGPDEQFLGRLMGAAAGIG 210

Search completed: May 6, 2004, 09:07:28
Job time: 26.7628 secs

RESULT 15
Q8E2E9 PRELIMINARY; PRT: 305 AA.
AC Q8E2E9;
DT 01-MAR-2003 (TrEMBLrel. 21, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylneuraminate lyase, putative.
GN SAG0039.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AE014193; AM98947.1; -;
DR TIGR; SAG0039; -;
DR GO; GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002220; DHDPS.
DR InterPro; IPR000437; Prok_lipoprot_s.
DR Pfam; PF00701; DHDPS; 1.
DR PRINTS; PRO0146; DHPICSNTHASE.

DR ProDom; PD001859; DHDPS; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW lyase; Complete proteome.
SQ SEQUENCE 305 AA; 33479 MW; 89E062907C1AF615 CRC64;

Query Match 21.3%; Score 247; DB 16; Length 305;
Best Local Similarity 32.1%; Pred. No. 7.4e-13;
Matches 68; Conservative 40; Mismatches 98; Indels 6; Gaps 6;

QY 6 XKQGLVAAATITPMTENGSEINFSVIGQYVYLVKEQGVKNIFVNGTGTGEGLSLSVSERRQ 65
DB 5 QKVGIIIPAFYACYDDKGDICPERVKALTYFT-DKGVQGLYVNGSSGECIYQSVADKRL 63
QY 66 VAEWVTKGKDKLDQVTHVHGALSLSKESQELACHAAEIGADGIVIAPIPFELKPTWTKDILL 125
DB 64 VLENVMSVAKGKL-TVIAHVACNNTKDSVELAMHAEIGVDAIAALPTIYFR-LPEYALIA 121
QY 126 NFLKEVAAAPALPFYVYHIPALTVGKIRAEELDGLDKIPTFGQKFSDTDLDFGQC 185
DB 122 DYNTISQAAPQDTFIYINIPQAGVAL-TSDLYRKMLCN-PQVIGVKNSSMPVQDIQNF 179
QY 186 VQNRQQQFAFLFGVDEQLLSALVMGATGAVG 217
DB 180 VAIGGENHIVF-NGPDEQFLGRLMGAAAGIG 210

Search completed: May 6, 2004, 09:07:28
Job time: 26.7628 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:54 ; Search time 9.57633 Seconds
(without alignments)
1239.929 Million cell updates/sec

Title: US-09-930-440B-2
Perfect score: 1.59
Sequence: 1 MAFFPKKQLQGLVAATITPTMTENGEINFSV-GQYVDYLKVGQVKNIFVNGTTGGLSLSV 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2.6/prodata/2/iaa/5A COMB.pap.*
2: /cgn2.6/prodata/2/iaa/5B COMB.pap.*
3: /cgn2.6/prodata/2/iaa/6A COMB.pap.*
4: /cgn2.6/prodata/2/iaa/6B COMB.pap.*
5: /cgn2.6/prodata/2/iaa/ECTOS COMB.pap.*
6: /cgn2.6/prodata/2/iaa/backfiles1.pap.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1159	100.0	230	US-09-516-143A-6	Sequence 6, Appli
2	1104.5	95.3	320	US-09-027-013-1	Sequence 1, Appli
3	1104.5	95.3	320	US-09-244-233-1	Sequence 1, Appli
4	394	34.0	300	US-09-543-681A-6868	Sequence 6868, Ap
5	276.5	23.9	297	US-09-027-013-3	Sequence 3, Appli
6	276.5	23.9	297	US-09-244-233-3	Sequence 3, Appli
7	269.5	23.3	308	US-09-543-681A-4426	Sequence 4426, Ap
8	248.5	21.4	318	US-09-027-013-4	Sequence 4, Appli
9	248.5	21.4	318	US-09-244-233-4	Sequence 4, Appli
10	241.5	20.8	341	US-09-328-352-5092	Sequence 5092, Ap
11	239.5	20.7	302	US-09-107-532A-5280	Sequence 5280, Ap
12	222.5	19.2	301	US-08-674-168-31	Sequence 31, Appl
13	222.5	19.2	301	US-08-985-928-15	Sequence 15, Appl
14	222.5	19.2	301	US-08-852-730-20	Sequence 20, Appl
15	220	19.0	294	US-09-489-039A-9498	Sequence 9498, Ap
16	214	18.5	290	US-09-631-828A-2	Sequence 2, Appli
17	207.5	17.9	299	US-09-134-001C-5221	Sequence 5221, Ap
18	196.5	17.0	299	US-09-252-991A-29593	Sequence 29593, A
19	196.5	17.0	318	US-09-540-226-1976	Sequence 1976, Ap
20	177	15.3	297	US-09-328-352-5386	Sequence 5386, Ap
21	170	14.7	292	US-08-973-461A-4	Sequence 4, Appli
22	170	14.7	292	US-08-648-010-4	Sequence 4, Appli
23	169.5	14.6	301	US-09-134-000C-5174	Sequence 5174, Ap
24	165	14.2	295	US-09-489-039A-9165	Sequence 9165, Ap
25	162	14.0	334	US-09-252-991A-20593	Sequence 20599, A
26	158	13.6	253	US-09-724-623-111	Sequence 111, App
27	158	13.6	314	US-09-328-352-6781	Sequence 6781, Ap

ALIGNMENTS

RESULT 1

US-09-516-143A-6
; Sequence 6, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PFS05PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; CURRENT FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-516-143A-6
Query Match 100.0%; Score 1159; DB 4; Length 230;
Best Local Similarity 100.0%; Pred No. 2.5e-225;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAFPKKKQLQGLVAATITPTMTENGEINFSV-GQYVDYLKVGQVKNIFVNGTTGGLSLSV	60
DB	1	MAFPKKKQLQGLVAATITPTMTENGEINFSV-GQYVDYLKVGQVKNIFVNGTTGGLSLSV	60
QY	61	SRRQVAEEVWTGKDKLDQVITHVGALSLSKESQELACHAAEIGADGLAVIAPFFLEWT	120
DB	61	SRRQVAEEVWTGKDKLDQVITHVGALSLSKESQELACHAAEIGADGLAVIAPFFLEWT	120
QY	121	KDILINFLKVAAPALPFYVYHIPALTVGKIPABELLGDLKIPFQGLKFSDDL	180
DB	121	KDILINFLKVAAPALPFYVYHIPALTVGKIPABELLGDLKIPFQGLKFSDDL	180
QY	181	DFGQCVQNRQQQPAFLFGVDEQLLSALVMGATGAVGFSVSRDLSTLLSN	230
DB	181	DFGQCVQNRQQQPAFLFGVDEQLLSALVMGATGAVGFSVSRDLSTLLSN	230

RESULT 2

US-09-027-013-1
; Sequence 1, Application US/09027013
; Patent No. 5962302
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: PCs
SOFTWARE: FASTSQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/C3/027,013
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0462 US
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4186
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT07
CLONE: 2125658
US-09-027-013-1

```

Query Match          95.3%;   Score 1104.5;   DB 3;   Length 320;
Best local Similarity 91.0%;   Pred.No.8.le-119;
Matches 222; Conservative    2; Mismatches    5; Indels    15; Gaps    1;

QY      1 MAFFPKKQLGLVAATITPMTENGEINFSVIGQVDVLVKEQGVKNIFVANGTTGEGLSLV 60
DBD     |||
QY      1 MAFFPKKKQLGLVAATITPMTENGEINFSVIGQVDVLVKEQGVKNIFVANGTTGEGLSLV 60
DBD     |||
        61 SRRQVABEWWTKGDKLDQVIHHVGALSKESQELAQHAAEIGADGIAVIAFPFLKPWT 120
        |||
        61 SRRQVABEWWTKGDKLDQVIHHVGALSKESQELAQHAAEIGADGIAVIAFPFLKPWT 120
DBD     |||
        121 KDILNLFKEVAAPALPFYYHYHPALTGVKIRAEELDGILDKIPTFOGLXFSDDTL 180
        |||
        121 KDILNLFKEVAAPALPFYYHYHPALTGVKIRAEELDGILDKIPTFOGLAFSDDTL 180
DBD     |||
        181 DFGQCVDQRQQOFAFLFGVDQLLASLWVGATGAVGS-----FVSDDL 225
        |||
        181 DFGQCVDQRQQOFAFLFGVDQLLASLWVGATGAVGSTVNYLGKKTQMLEAFEQKDPS 240
        |||
        226 TLLS 229
        | :
        241 LAIN 244

RESULT 4
US-09-543-681A-6868
; Sequence 6868, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEU
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681-A
; CURRENT FILING DATE: 2000-04-05
```

RESULT 4
US-09-543-681A-6868
? Sequence 6868, Application US/09543681A
? Patent No. 6605709
? GENERAL INFORMATION:
? APPLICANT: GARY BRETON
? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI
? TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
? FILE REFERENCE: 2709.1002-001
? CURRENT APPLICATION NUMBER: US/09/543,681A
? CURRENT FILING DATE: 2000-04-05

```

; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6868
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6868

Query Match
Best Local Similarity 34.0%; Score 394; DB 4; Length 300;
Matches 87; Conservativity 39; Mismatches 82; Indels 4; Gaps 4;

QY 7 KLGGLVAATITPMTENGESVIGQVYVYLVKQGVKNI FVNGTTGEGLSLSVSERRQV 66
Db 6 KLSGLIAAHTFAADGSVNPVIDDIAKELI-TTGTGAYVLGTTGEGHCSVEERKV 64
QY 67 AEWVTGKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFFLKPTWKDILIN 126
Db 65 AERWVTASQGLD-LIHTGALSIADTLELARAHAETLDIKATSVIGPCFFKFSHVDDLVE 123
QY 127 ELKEVAAAAPALPFYVYHPIALTCVKIRAEELDGILDKIPTFQGLKFSDDLDDFGQCV 186
Db 124 YCLAAASAPSFGFYH-STWSGLSDMEKFLQAAKGVJENUSGKFNSPDWEFORCL 182
QY 187 DQNRQOQFAFLFGVDRQLLSALVWGATGAVGS 218
Db 183 -RVEGKYLIPFGVDFEIPAGLACGALSAGVS 213

```

```

RESULT 5
US-09-027-013-3
; Sequence 3, Application US/09027013
; Patent No. 5963302
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,013
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0462 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: GI 42131
US-09-244-233-3

```

```

; LIBRARY: GenBank
; CLONE: GI 42131
US-09-027-013-3

Query Match
Best Local Similarity 23.9%; Score 276.5; DB 2; Length 297;
Matches 59; Conservativity 48; Mismatches 86; Indels 9; Gaps 6;

QY 8 LQGLVAATITPMTENGESVIGQVYVYLVKQGVKNI FVNGTTGEGLSLSVSERRQVA 67
Db 5 LRGWMAALLTPDQOQALDKASLRRLVQFNI-QQGDGLYVGGSTGEAFVQSLSEREQVL 63
QY 68 EEWVTGKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFFLKPTWKDILIN 127
Db 64 EIVAEKGGKI-KLIAHVGCVITRESQQLAASAKRYGFDASVATPPY-PF3FEHCDH 121
QY 128 LKEVAAAAPALPFYVYHPIALTCVKIRAEELDGILDKIPTFQGLKFSDDLDDFGQCV 187
Db 122 YRAIDSADGLPMVYVNIIPALSGVLTLDQI--NTLVLPVGALKQTSGLYQMEQI-- 177
QY 188 QNRQOQFAFLFGVDRQLLSALVWGATGAVGS 218
Db 178 -RREHDELVYNGYDEIFASGLLAGADGGIGS 208

```

```

RESULT 6
US-09-244-233-3
; Sequence 3, Application US/09244233
; Patent No. 6030824
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,233
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/027,013
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0462 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: GI 42131
US-09-244-233-3
Query Match
23.9%; Score 276.5; DB 3; Length 297;

```

RESULT 8
US-09-027-013-4
; Sequence 4, Application US/39027013
; Patent No. 5952302
; GENERAL INFORMATION:
; APPLICANT: Hillmar, Jennifer L.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN N-ACETYLNEURAMINATE LYASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Invivo Pharmaceuticals, Inc.


```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,233
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 39/027,013
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0462 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: GI 1016806
; US-09-244-233-4
;
; Query Match 21.4%; Score 248.5; DB 3; Length 318;
; Best Local Similarity 30.0%; Pred. No. 4.4e-20;
; Matches 64; Conservative 47; Mismatches 95; Indels 7; Gaps 5;
;
QY 6 KXQLGLVAATITPTMTENGINSVIGQYVDYLVYKEQGVKNIFVNGTGTGEGLSLSVSRQYA 65
DB 27 KSLKGLFSALLSFNEDGTINEKGLREIVRYNIDKMKIDGLYVGGSTGENFELSTEEKQ 86
QY 66 VAEQEWTKGKLDQVLIHVGLSLKESQELAHAAEIGADGAVIAPFPFLKWTKDILI 125
DB 87 IFRANDKAOQV-ALIAQVGSINIHESIELGKIATGELGNCLSAVTPFYK-FTFEIK 144
QY 126 NELKEVAAAPALPFYYPHYIPALTGVKIRAEELDGLDKIPTFQGLKFSDDTLDFGQC 185
DB 145 NYNIVNAT-GNMIVYISIPALTGVSMADQF-GELFENPKIIGVKFTAGFYLLERV 201
QY 186 VQNFQOQAFPLFGVDEQLLSALVMGATGAVS 218
DB 202 --KRAYPDHLIWAQFDEMMLPACSLGIDGAIGS 232
;
; RESULT 10
; US-09-328-352-5092
; Sequence 5092, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5092
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-5092
;
; Query Match 20.8%; Score 241.5; DB 4; Length 341;
; Best Local Similarity 30.6%; Pred. No. 3.1e-19;
; Matches 64; Conservative 49; Mismatches 89; Indels 7; Gaps 6;
;
QY 8 LQGLVAATITPTMTENGINSVIGQYVDYLVYKEQGVKNIFVNGTGTGEGLSLSVSRQYA 67
DB 52 IQGSIVAVITPMLKDGVDKSLKLEKLEWHI-EQGTNSIVAVGTTGEASTLSMEETQVI 110
QY 68 BEWTKGKLDQVLIHVGLSLKESQELAHAAEIGADGAVIAPFPFLKWTKDILIINF 127
DB 111 KE-IIRVANKRIPILNGTGANSTRAIELTKAAGADAAALLVTEYVYK-TOEGLYOH 168
QY 128 LKEVAAAPALPFYYPHYIPALTGVKIRAEELDGLDKIPTFQGLKFSDDTLDFGQCVD 187
DB 169 YKAIAEAV-ELFLILYVNGRTGVDSNDTAVR-LAEIPNIVGIKDATGD-VPRGKALI 224
QY 188 QNRQOQAFPLFGVDEQLLSALVMGATGAV 216
DB 225 DALNGKMAVYSGDDETAMELMILGADGNI 253
;
; RESULT 11
; US-09-107-532A-5280
; Sequence 5280, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <UNKNOWN>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,499
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5280:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...302
; SEQUENCE DESCRIPTION: SEQ ID NO: 5280:
; US-09-107-532A-5280
;
; Query Match 20.7%; Score 239.5; DB 4; Length 302;
; Best Local Similarity 29.4%; Pred. No. 4.4e-19;
; Matches 63; Conservative 51; Mismatches 87; Indels 13; Gaps 7;
;
QY 6 KXQLGLVAATITPTMTENGINSVIGQYVDYLVYKEQGVKNIFVNGTGTGEGLSLSVSRQ 65

```

Db 9 KMKIGSIVALTIPMNEEGSDVAGLEKLIQFHDEQ--TGGLVLGTGSSSTLTQSEEEQ 67
QY 66 VABEWTKGKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFLPKWTKDILI 125
Db 68 ILIOLTVKRVAGRV-PVIAGAGTNTKETTIEKAKHFA--GADALVITPYNKTSAGLAA 126
QY 126 NFKUVEAAMAPALPFYVYHIPALTCVKIRAEELDGLDKIPTFQGLKFSDDTL--LDF 182
Db 127 HF-TAARESP-IPLIILYNVPSRTGMSIPTHLVN--LAEHPNIGLKEASGDMAYWMDA 182
QY 183 GQCVQDQROQOAFPLFGVDEQLLSALVMGATGAV 216
Db 183 ARLIG----EEFTLYSGNDLILPVMVSGGSGVI 212

RESULT 12
US-08-674-168-31
; Sequence 31, Application US/08674168
; Patent No. 5804414
; GENERAL INFORMATION:
; APPLICANT: MORIYA, Mika
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: YOKOZUKI, Kenzo
; APPLICANT: HIRANO, Seiko
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: IZUI, Masako
; APPLICANT: SUGIMOTO, Masakazu
; TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
; TITLE OF INVENTION: ARTIFICIAL TRANSPOSON
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,168
; FILING DATE: 01-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-166541
; FILING DATE: 30-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-674-168-31

Query Match 19.2%; Score 222.5; DB 1; Length 301;
Best Local Similarity 29.6%; Pred. No. 4e-17;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;
QY 14 ATITPMTENGINFVIGQYVDYLVKEQGVKNIFVNGTTGEGLSLSVSRQVAEHWTK 73

Db 20 AMVTFETSGDIDIAAGREVAAYLV-DKGLDSLVLACTTGESPTTAAEKELELKAVREE 78
QY 74 GKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFLPKWTKDILINFLKXVAA 133
Db 79 VGDRAN-VIAGVGTNTTTSVELAAEASAGADGLLVVTPYYSKPSQEGLLAHP--GAIA 135
QY 134 AAPALPFYVYHIPALTCVKIRAEELDGLDKIPTFQGLKFSDDTLDFGQCVQDQNEQQQ 193
Db 136 AATEVPCILYDIPGRSGIPESDTMR--LSELPITLAVKDAKGLVAATSLI---KETG 190
QY 194 FAELFGVDEQLLSALVMGATGAV 216
Db 191 LAWVGDDPLNLVWALGGSGFI 213

RESULT 13
US-08-985-908-15
; Sequence 15, Application US/08985908
; Patent No. 6024773
; GENERAL INFORMATION:
; APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSIYOSHI NA
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,908
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-325659
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-985-908-15

Query Match 19.2%; Score 222.5; DB 3; Length 301;
Best Local Similarity 29.6%; Pred. No. 4e-17;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;
QY 14 ATITPMTENGINFVIGQYVDYLVKEQGVKNIFVNGTTGEGLSLSVSRQVAEHWTK 73
Db 20 AMVTFETSGDIDIAAGREVAAYLV-DKGLDSLVLACTTGESPTTAAEKELELKAVREE 78
QY 74 GKDKLDQVLIHVGALSLSKESQELAQHAAEIGADGIAVIAPFLPKWTKDILINFLKXVAA 133
Db 79 VGDRAN-VIAGVGTNTTTSVELAAEASAGADGLLVVTPYYSKPSQEGLLAHP--GAIA 135
QY 134 AAPALPFYVYHIPALTCVKIRAEELDGLDKIPTFQGLKFSDDTLDFGQCVQDQNEQQQ 193
Db 136 AATEVPCILYDIPGRSGIPESDTMR--LSELPITLAVKDAKGLVAATSLI---KETG 190
QY 194 FAELFGVDEQLLSALVMGATGAV 216

us-09-930-440b-2.ra1

Thu May 13 11:52:59 2004

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 9498
 ; LENGTH: 294
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-3498

Query Match 19.0%; Score 220; DB 4; Length 294;
 Best Local Similarity 28.8%; Pred. No. 7.4e-17;
 Matches 65; Conservative 40; Mismatches 97; Indels 24; Gaps 8;
 QY 5 KKKLOGLVAATITMTENGESVIGVGVYDYLKKEQGVKN-IFVNGITGEGLSLSVSR 63
 DB 2 KRAISGVLTALVTPPTAGALNLPALRQVQ---PQLAAGNIGFCGGTNGEFFFVLEEEK 58
 QY 64 EQVABEWTGKDKLDQVTHVGALESXESQELAQHAAIGADGIAVIAPFLKPTKDI 123
 DB 59 LAVARTCVBAAAGRA-PVVAHIGEVSTRETRRLGQOLARLGVDVAISATEWV-PLKQEE 116
 QY 124 LINFLEKVAAPALPFYHYHIPALTGVKIRAE-----ELLGILDKIPTFOGLKFS 175
 DB 117 LINHYTAADAL-SVPEFLYNIPARTNTIAPETARQARHENIVGIKDSAGSIDSLK-- 173
 QY 176 DTDLDFGQCVQDNRQQQFAFLFGVDEQLLSALVMGATGAVGSFVS 221
 DB 174 ---GELDAVRDID-----GFDVLNGPDSLHQGFVDGCSACISGLAN 212

Search completed: May 6, 2004, 09:08:35
 Job time: 10.5763 secs

Db 191 LAWYSGDDPLNLVWLAGSGGFI 213
 ;

RESULT 14
 US-08-852-730-20
 ; Sequence 20, Application US/08852730
 ; Patent No. 6096597

; GENERAL INFORMATION:
 ; APPLICANT: SEIKO HIRANO, MASAKAZU SUGIMOTO, EICHI NAKANO,
 ; APPLICANT: MASAKO IZUI, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI
 ; APPLICANT: NAKAMATSU

; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE

; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MALER AND NEUSTADT, P.C.
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
 ; CITY: ARLINGTON
 ; STATE: VA
 ; ZIP: 22206

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/852,730

; FILING DATE: 05-07-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-142812

; FILING DATE: 05-JUN-1996

; ATTORNEY/AGENT INFORMATION:
 ; NAME: NORRAN F. OBLON

; REGISTRATION NUMBER: 24,618

; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 301 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLSCULE TYPE: protein

; US-08-852-730-20

Query Match 19.2%; Score 222.5; DB 3; Length 301;
 Best Local Similarity 29.6%; Pred. No. 4e-17;
 Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATITMTENGESVIGVGVYDYLKKEQGVKNIFVNGITGEGLSLSVSRQVAAEWVTK 73
 DB 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLVAGTTGESPTTAAEKLELLKAVREE 78
 QY 74 GKDKLDQVTHVGALESXESQELAQHAAIGADGIAVIAPFLKPTKDIILNFKVEVA 133
 DB 79 VGDRAN-VIAGVGTNNRTSVELAAASAGADGILLVTPVYKSPQEGGLAHF--GAIA 135
 QY 134 AAPALPFYHYHIPALTGVKIRAEELDGLDKPTFGGLKFSDTDLDFGQCVQDNRQQ 193
 DB 136 AATEVPICLYDIPRSGIPESDTMR--LSELPTILAVKADGDLVAATSLI---KETG 190
 QY 194 FAFLEGVDEQLLSALVMGATGAV 216
 DB 191 LAWYSGDDPLNLVWLAGSGGFI 213

RESULT 15
 US-09-489-039A-9498
 ; Sequence 9498, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al

Sequence 85, Appli
Sequence 45588, A
Sequence 16546, A
Sequence 43119, A
Sequence 23683, A
Sequence 13888, A
Sequence 56432, A
Sequence 716, Appli
Sequence 2, Appli
Sequence 21028, A
Sequence 55465, A
Sequence 19025, A
Sequence 18247, A
Sequence 14458, A
Sequence 4914, A
Sequence 17315, A
Sequence 17315, A
Sequence 336, Appl
Sequence 23132, A
Sequence 20676, A
Sequence 21450, A
Sequence 4687, Appl
Sequence 55, Appli
Sequence 2859, Appl
Sequence 46924, A
Sequence 7446, Appl
Sequence 19626, A
Sequence 17250, A
Sequence 9753, Appl
Sequence 20274, A

US-10-369-493-85
US-10-282-122A-45588
US-10-369-493-16546
US-10-282-122A-43119
US-10-369-493-23683
US-10-369-493-13888
US-10-282-122A-56432
US-10-369-493-716
US-10-214-556-2
US-10-369-493-21028
US-10-282-122A-55465
US-10-369-493-19025
US-10-369-493-18247
US-10-369-493-14458
US-10-369-493-4914
US-10-369-493-17315
US-10-369-493-17315
US-10-238-075-336
US-10-369-493-23132
US-10-369-493-20676
US-10-369-493-21450
US-10-369-493-4687
US-09-971-536-55
US-10-369-493-2859
US-10-282-122A-46924
US-10-369-493-7446
US-10-369-493-19626
US-10-369-493-17250
US-10-369-493-9753
US-10-369-493-20274

ALIGNMENTS

RESULT 1
US-09-984-205-6
; Sequence 6, Application US/09984205
; Patent No. US20020137175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: P5505D1
; CURRENT APPLICATION NUMBER: US/09/984,205
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-205-6

Query Match 100.0%; Score 1159; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.5e-119;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFPKKK:QGLVAATITPTMTNGEINFVSIGQYDYLKKEQGVKIFVNGTTEGGLSLV 60
Db 1 MAFPKKK:QGLVAATITPTMTNGEINFVSIGQYDYLKKEQGVKIFVNGTTEGGLSLV 60
QY 61 SRRROVAEWYTKGKLDQVLIHVGALSLKESQELAGDGLIAPFFPKPT 120
Db 61 SRRROVAEWYTKGKLDQVLIHVGALSLKESQELAGDGLIAPFFPKPT 120
QY 121 KDILNFKVAAPALPFVYVHIPALTGKIRAEELDGLIDKIPFGKFSFDLL 180

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:07:45 ; Search time 24.8649 Seconds
(without alignments)
2567.492 Million cell updates/sec

Title: US-09-930-440B-2
Perfect score: 1159
Sequence: 1 MAFPKKK:QGLVAATITPTMT.....GATGAVGSFVRDLSTLLEN 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1159	100.0	230	9	US-09-984-205-6
2	1159	100.0	230	9	US-09-930-440B-2
3	274.5	23.7	297	15	US-10-369-493-23570
4	274.5	23.5	297	14	US-10-182-504-14
5	251.5	21.7	294	15	US-10-369-493-21648
6	241	20.8	313	15	US-10-369-493-18552
7	234.5	20.2	299	15	US-10-369-493-8308
8	231.5	20.0	283	15	US-10-369-493-1171
9	230.5	19.9	287	15	US-10-369-493-1265
10	230.5	19.9	287	15	US-10-369-493-20331
11	226	19.5	290	15	US-10-369-493-11339
12	222.5	19.2	301	7	US-08-952-976-15
13	222.5	19.2	301	9	US-09-738-626-6976
14	222.5	19.2	301	10	US-09-746-650A-54
15	222.5	19.2	301	14	US-10-226-136-15

```

; LENGTH: 297
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-23570

Query Match
  23.7%; Score 274.5; DB 15; Length 297;
Best Local Similarity 34.3%; Pred. No. 2.1e-21;
Matches 74; Conservative 43; Mismatches 82; Indels 17; Gaps 8;

QY 8 LOGLVAAITPTMTENGELNFSVIGQYVDYLKKEQVKNIFVNGTGTGEGLSLSVSRRQ-- 65
DB 5 LRGVMAALLTPDQQQALDKASRLRLVQNT--QQGIDGLYVGGSTGEAFVQSLSREQLV 63
QY 66 --VAEEWTKGDKLDQVLIHVGALSLSKESQELAAHAEIGADGIAVIAPFFLKWTXDI 123
DB 64 EIVAE--ARKGKIL--IAHVCVSTAESQQLASAKRYGFDVAVTPFFY-PFSEEE 117
QY 124 LINFELKEVAAAAPALPFYHYHIPALTVGKIRAEELLDGILDKIPTFQGLKFSDTLLD 183
DB 118 HCDHYRAIIDSADGLPMVYVNIIPALSGVKLTLDQI--NTLVLPVGVALKQTSGLYQME 175
QY 184 QVDONROQQAFLF-GYDEQLLSALVMGATGAVGS 218
DB 176 QT---RREHPDLVLYNGYDEIFPAGLLAGADGGIGS 208

RESULT 4
US-10-182-504-14
; Sequence 14, Application US/10182504
; Publication No. US20030157121A1
; GENERAL INFORMATION:
; APPLICANT: Chatfield, Steven Neville
; TITLE OF INVENTION: Virulence Genes, Proteins, and Their Use
; FILE REFERENCE: GJE-98
; CURRENT APPLICATION NUMBER: US/10182,504
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: PCT/GB01/00449
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: GB 0023059.9
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: GB 0023060.7
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: GB 0002552.8
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: GB 0002554.4
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: GB 0002555.1
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: GB 0002556.9
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: GB 0002557.7
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: GB 0002558.5
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: GB 0002559.3
; PRIOR FILING DATE: 2000-02-03
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Salmonella typhimurium
US-10-182-504-14

Query Match
  23.5%; Score 272.5; DB 14; Length 297;
Best Local Similarity 31.9%; Pred. No. 3.6e-21;
Matches 68; Conservative 43; Mismatches 95; Indels 7; Gaps 5;

QY 6 KKLQGLVAATITMTENGELNFSVIGQYVDYLKKEQVKNIFVNGTGTGEGLSLSVSRRQ 65
DB 3 KALQGVMAALLTPDQQQALDKASRLRLVQNT--QQGIDGLYVGGSTGEAFVQSLAEREQ 61
```

```

; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-930-440B-2

Query Match
  100.0%; Score 1159; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 8.5e-119;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFPKKXQLGLVAATITPTMTENGELNFSVIGQYVDYLKKEQVKNIFVNGTGTGEGLSLV 60
DB 1 MAFPKKXQLGLVAATITPTMTENGELNFSVIGQYVDYLKKEQVKNIFVNGTGTGEGLSLV 60
QY 61 SRRQVAEEWTKGDKLDQVLIHVGALSLSKESQELAAHAEIGADGIAVIAPFFLKWTX 120
DB 61 SRRQVAEEWTKGDKLDQVLIHVGALSLSKESQELAAHAEIGADGIAVIAPFFLKWTX 120
QY 121 KDILINFLKEVAAAAPALPFYHYHIPALTVGKIRAEELLDGILDKIPTFQGLKFSDTLL 180
DB 121 KDILINFLKEVAAAAPALPFYHYHIPALTVGKIRAEELLDGILDKIPTFQGLKFSDTLL 180
QY 181 DFGQCVQDNRQQAFLFVGDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230
DB 181 DFGQCVQDNRQQAFLFVGDEQLLSALVMGATGAVGSFVSRDLSTLLSN 230

RESULT 3
US-10-369-493-23570
; Sequence 23570, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23570
```

Thu May 13 11:53:00 2004

```

QY 66 VAEWVTGKDKLDQVLIHVGAISLKSQELAQHAAEAGAGIAVIAPFFFLKPTWKDILI 125
Db 62 VLEIWAESAAGKI-TLIAHVGTSTVAESQQLASAAKRYGFDVAISAVTPFYI-PFSFEHHC 119
QY 126 NFLKEVAAAAPALPFYHYHHPALTVGKIRABELLGILDKIPTFQGLKESDTLLDFGQC 185
Db 120 DHYRAIIDSADGLPFWVYNIFALSGVKLTLDQI--NTLVTLFGVSALKQTSGLDFQMEQI 177
QY 186 VDNRRQOQFAFLFGVDEQLLSALVMGATGAVS 218
Db 178 --RRAHPDLVLYNGIDEIFASGLLAGAGGIGS 208

RESULT 5
US-10-369-493-21648
; Sequence 21648, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21648
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21648

Query Match 21.7%; Score 251.5; DB 15; Length 294;
Best Local Similarity 30.1%; Pred. No. 7.1e-19;
Matches 65; Conservative 49; Mismatches 81; Indels 21; Gaps 7;

QY 8 LQGLVAATITPMTENGINSFVIGQVVDYLVKQGVKNIFVNGTGTGEGLSLSVSERQVA 67
Db 1 MEGVIVPLVTPFREDYSIDFEALEWHIRFL-EEGVHGIFVNSTGTGFTSLNTERKMLA 59
QY 68 EEW-VTKGDKLDQVLIHVGAISLKSQELAQHAAEAGAGIAVIAPFFFLKPTWKDILI 125
Db 60 EKGREITSG-----MYLVGTGSTSTLEVIESRRAEDIGADGIVIVAPYYCKLXDEILK 114
QY 126 NFLKEVAAAAPALPFYHYHHPALTVGKIRABELLGILDKI-----PTFQGLKFSOTDILLD 181
Db 115 HF--SWAERVVDIPIIIVVAIPSC-----ANPVPVDIIRKVSLEYSNIIIGVKASVDSLTY 166
QY 182 FGQVCD-QNRQOQFAFLFGVDEQLLSALVMGATGAV 216
Db 167 LQELIEVKEERKDPRTVFTGLDQYFLSTILLJGDDGI 202

RESULT 6
US-10-369-493-18552
; Sequence 18552, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

```

```

; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18552
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-18552

Query Match 20.9%; Score 241; DB 15; Length 313;
Best Local Similarity 33.6%; Pred. No. 1.1e-17;
Matches 72; Conservative 33; Mismatches 95; Indels 14; Gaps 7;

QY 8 LQGLVAATITPMTENGINSFVIGQVVDYLVKQGVKNIFVNGTGTGEGLSLSVSERQVA 67
Db 25 LRGWVPVVTAFDADESLEADTTADHA-RMVDAGVHGVPLGTNGEFPFLITSERDRV- 82
QY 68 EEWVTGKDKLDQVLIHVGAISLKSQELAQHAAEAGAGIAVIAPFFFLKPTWKDILI 124
Db 83 ---VTAVDEVGGEVFIAGVGAPSTRQTVAAHAAHAAVSGADGVVWVTPFYI-PLDGTAA 138
QY 125 INFLKEVAAAAPALPFYHYHHPALTVGKIRABELLGILDKIPTFQGLKESDTLLDFGQ 184
Db 139 VEHYRVAADVDC-PVYVTHIPSKTGNLSLETL--AALAEIPTLAGVKSXKDVPMWIGQ 195
QY 185 CYDQNRQOQFAFLFGVDEQLLSALVMGATGAVS 218
Db 196 AVDAH--PELTFLAGSDSLLAPGLDVGCAGLVSA 227

RESULT 7
US-10-369-493-8308
; Sequence 8308, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8308
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-369-493-8308

Query Match 20.2%; Score 234.5; DB 15; Length 299;
Best Local Similarity 29.8%; Pred. No. 5.4e-17;
Matches 68; Conservative 39; Mismatches 110; Indels 11; Gaps 5;

QY 4 PKKKLQGLVAATITPMTENGINSFVIGQVVDYLVKQGVKNIFVNGTGTGEGLSLSVSER 63
Db 7 FNAFPGQMLTAMITPMLDNGEVDYGVARLATYLVDEQNRDGLIYVNGTGTGEGLSLSVSER 66
QY 64 QVAAEWTGKDKLDQVLIHVGAISLKSQELAQHAAEAGAGIAVIAPFFFLKPTWKDILI 123
Db 67 BRILRTVIDAVGDRA-TIVAGAGSNDTRES-ELAKTAERAGAGDGLLVTPYINRPPQGL 125
QY 124 INFLKEVAAAAPALPFYHYHHPALTVGKIRABELLGILDKIPTFQGLKESDTLLDFGQ 183
Db 126 LRHP--THAIDATGLPIMLYDIPGTGTPIDETLVR--LAHPRIIVAKDAKDDL---G 178
QY 184 QCVQDNRQOQFAFLFGVDEQLLSALVMGATG---AVGSFVSRLSTLL 228
Db 179 ASSWVMSRTDLAYSGSDMLNPLLSIGAGAFVSVGVHVGVSSELHMI 226

```


us-09-930-440b-2.rapb

Thu May 13 11:53:00 2004

APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 11339
LENGTH: 290
TYPE: PRT
ORGANISM: Methanosarcina mazei
US-10-369-493-11339

Query Match 19.5%; Score 226; DB 15; Length 290;
Best Local Similarity 29.3%; Pred. No. 4,5e-16;
Matches 61; Conservative 45; Mismatches 94; Indels 8; Gaps 4;

QY 9 QGLVAATTPMTENGESINFSVIGQYVDYLVKQGVKNIFVNGTTGEGLSVSRQVAE 68
DB 3 EGAMFALITFTKDRIDREGLENIET-VERGGVAGIVPGCTTGESATLSALEHEVID 61
QY 69 EWTGKOKLDQVIHVGALSLSKESQELAQHAAGIAGDIAVIAPFFFLKPTWKDILINFL 128
DB 62 IAVECSK---VPVIAGTGSNNTGEALQFTKHAADAGVGVLLISPYNNKPNAGLLTHFK 118
QY 129 KEVAAAPALPFYVYHIPALTVKIRAEELDGILDKIPTFQGLKPSDTLLDFGQCVQD 188
DB 119 K--IAEAVDVPWVNNIPBRTGCDMPLEIVE--LAKVENIVIGKEASGNIGKVSQILEN 174
QY 189 NRQQQFAFLFGVDEQLLSALVMGATGAV 216
DB 175 TIDEDFVVISGEDNLTLPLSVGGQGV 202

RESULT 12
US-08-952-976-15
Sequence 15, Application US/08952976
Publication No. US20020086370A1
GENERAL INFORMATION:
APPLICANT: QTSUNA, Seiko
APPLICANT: SUGIMOTO, Masakazu
APPLICANT: IZUI, Masako
APPLICANT: HAYAKAWA, Atsushi
APPLICANT: NAKANO, Eiichi
APPLICANT: KOBAYASHI, Masaki
APPLICANT: YOSHIMURA, Yasuhiko
APPLICANT: NAKAMATSU, Tsuyoshi
TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,976
FILING DATE: 8-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-140614
FILING DATE: 07-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618

TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 301 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-952-976-15

Query Match 19.2%; Score 222.5; DB 7; Length 301;
Best Local Similarity 29.6%; Pred. No. 1.1e-15;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATTPMTENGESINFSVIGQYVDYLVKQGVKNIFVNGTTGEGLSVSRQVAEHWTK 73
DB 20 AMVTPFTESGDIDIAAGREVAAYLV-DKGLDSLVLVAGTTGESPTTAAAEKLELLKAVREE 78
QY 74 GKDKLDQVIHVGALSLSKESQELAQHAAGIAGDIAVIAPFFFLKPTWKDILINFLKEVAA 133
DB 79 VGRAN-VIAGVTNNTRTSVELAEAAAGADGLLVVTPYYSKPSQEGLLAHP--GAIA 135
QY 134 AAPALPFYVYHIPALTVKIRAEELDGILDKIPTFQGLKPSDTLLDFGQCVQDNRQQ 193
DB 136 ATEVPICLYDIPGRSSIPESDTMR--LSELTILAVKDAKGLVAATSLI---KETG 190
QY 194 FAPFGVDEQLLSALVMGATGAV 216
DB 191 LAWYSGDDPLNLVWLALGGSGFI 213

RESULT 13
US-09-738-626-6976
Sequence 6976, Application US/09738626
Publication No. US20020197635A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6976
LENGTH: 301
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6976

Query Match 19.2%; Score 222.5; DB 9; Length 301;
Best Local Similarity 30.0%; Pred. No. 1.1e-15;
Matches 62; Conservative 47; Mismatches 81; Indels 17; Gaps 6;

QY 14 ATTPMTENGESINFSVIGQYVDYLVKQGVKNIFVNGTTGEGLSVSRQVAEHWTK 69
DB 20 AMVTPFTESGDIDIAAGREVAAYLV-DKGLDSLVLVAGTTGESPTTAAAEKLELLKAVREE 78

Thu May 13 11:53:00 2004

```

QY 70 WTKGKDKLDQVLIHVGALSLKESQELAHQAEIGADG-AVIAPFFLKPKWTKDILINFLK 129
Db 79 VGDRAK-----LIAGVTNNTRTSVLAERAAAGAGGLVVT-PYYSKPSQEGLLAHF-- 131
QY 130 EVAAARALPFYHYHHPALTVGKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVQDN 189
Db 132 GAIAAATEVPICLYDIPGRSGIPESDTWR--LSELPILAVKADKGLVAATSLI--- 186
QY 190 RQOQFAFLFGVDEQLLSALVMGATGAV 216
Db 187 KETGLAWYSGDDFLNLVWLALGSGFI 213

```

RESULT 14

```

US-09-746-660A-54
; Sequence 54, Application US/39746662A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Woc
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: B01-121C2
; CURRENT APPLICATION NUMBER: US/09/746,660A
; CURRENT FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 54
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-746-660A-54

```

```

Query Match 19.2%; Score 222.5; DB 10; Length 301;
Best Local Similarity 30.0%; Pred. No. 1.1e-15;
Matches 62; Conservative 47; Mismatches 81; Indels 17; Gaps 6;

QY 14 ATITPMTENGINFVIGQVYDLVKEQGVKNIFVNGTGGELSLSVSR-----RQVARE 69
Db 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLGTTGSPPTTAACKLELLKAVREE 78
QY 70 WTKGKDKLDQVLIHVGALSLKESQELAHQAEIGADG-AVIAPFFLKPKWTKDILINFLK 129
Db 79 VGDRAK-----LIAGVTNNTRTSVLAERAAAGAGGLVVT-PYYSKPSQEGLLAHF-- 131
QY 130 EVAAARALPFYHYHHPALTVGKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVQDN 189
Db 132 GAIAAATEVPICLYDIPGRSGIPESDTWR--LSELPILAVKADKGLVAATSLI--- 186
QY 190 RQOQFAFLFGVDEQLLSALVMGATGAV 216
Db 187 KETGLAWYSGDDFLNLVWLALGSGFI 213

```

```

RESULT 15
US-10-226-136-15
; Sequence 15, Application US/10226136
; Publication No. US20030054506A1
; GENERAL INFORMATION:
; APPLICANT: OTSUNA, Seiko
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: IZUI, Masako
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: NAKANO, Eiichi
; APPLICANT: KOBAYASHI, Masaki
; APPLICANT: YOSHIHARA, Yasuhiko
; APPLICANT: NAKAMATSU, Tsuyoshi
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/226,136
; FILING DATE: 23-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,976
; FILING DATE: 8-DEC-1997
; APPLICATION NUMBER: JP 7-140614
; FILING DATE: 07-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-226-136-15

```

```

Query Match 19.2%; Score 222.5; DB 14; Length 301;
Best Local Similarity 29.6%; Pred. No. 1.1e-15;
Matches 60; Conservative 50; Mismatches 84; Indels 9; Gaps 5;

QY 14 ATITPMTENGINFVIGQVYDLVKEQGVKNIFVNGTGGELSLSVSR-----RQVARE 73
Db 20 AMVTFTESGDIDIAAGREVAAYLV-DKGLDSLVLGTTGSPPTTAACKLELLKAVREE 78
QY 74 GKDKLDQVLIHVGALSLKESQELAHQAEIGADG-AVIAPFFLKPKWTKDILINFLK 133
Db 79 VGDRAN-VIAGVTNNTRTSVLAERAAAGAGGLVVT-PYYSKPSQEGLLAHF--GAIA 135
QY 134 AAPALPFYHYHHPALTVGKIRAEELDGLDKIPFQGLKFSDDTLDDFGQCVQDN 193
Db 136 AATAATEVPICLYDIPGRSGIPESDTWR--LSELPILAVKADKGLVAATSLI---KETG 190
QY 194 FAFLFGVDEQLLSALVMGATGAV 216
Db 191 LAWYSGDDFLNLVWLALGSGFI 213

```

us-09-930-440b-2.rapb

Thu May 13 11:53:00 2004

Search completed: May 6, 2004, 09:25:04
Job time : 25.8649 secs

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:34 ; Search time 5296.32 Seconds
(without alignments)
10679.608 Million cell updates/sec

Title: US-09-930-440b-3
Perfect score: 1305
Sequence: 1 atggactcggggagaagg.....ataattcgcacaaatag 1305

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 6
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Result No.	Score	Query Match %	Length	DB ID	Description
1	1303.4	99.9	1305	6	BD270234 Human gly
2	1280.4	98.1	1305	9	AF397212 Homo sapi
3	1272.4	97.5	1376	9	AF271388 Homo sapi
4	1272.4	97.5	1766	9	HSMB34286
5	1272.4	97.5	1770	6	AR338869 Sequence
6	1270.8	97.4	1634	6	AX882419 Sequence
7	1270.8	97.4	1634	6	BD159714 Primer fo
8	1270.8	97.4	1694	9	AK022927 Homo sapi
9	1270.8	97.4	1772	6	BD260951 Carboxydr
10	1270.8	97.4	1772	6	AX040083 Sequence
11	1008.8	77.3	1704	10	BC063776 Mus muscu
12	1008.8	77.3	1728	10	BC031500 Mus muscu
13	1007.2	77.2	1687	10	MMU6215
14	918.4	70.4	1588	9	BC016609 Homo sapi
15	636.6	48.8	838	6	AX541147 Sequence
16	625	47.9	819	6	AX543862 Sequence
17	620.2	47.5	748	6	AX541885 Sequence
18	562	43.1	686	6	AX541885 Sequence
19	538.2	41.2	622	6	AX868002 Sequence
20	538.2	41.2	622	6	BD148064 Primer fo
21	399	30.6	889	10	BC004606 Mus muscu
22	359	27.5	1468	5	AB027414 Oncorhync
23	260	19.9	625	9	HS332638 Homo sapi
24	260	19.9	72001	9	AC092862 Homo sapi
25	260	19.9	151221	2	AC093092 Homo sapi
26	260	19.9	153201	2	AC092863 Homo sapi
27	260	19.9	175371	2	AC092822 Homo sapi
28	252	19.3	729	9	HS3330107 Homo sapi
29	215	16.5	203003	2	AC087313 Homo sapi
30	184.4	14.1	2136	9	AK124990 Homo sapi
31	184.4	14.1	153201	2	AC092863 Homo sapi
32	184.4	14.1	156314	9	AX873073 Sequence
33	181.6	13.9	589	6	BD153135 Primer fo
34	181.6	13.9	589	6	BD153135 Primer fo
35	177.4	13.6	151221	2	AC093092 Homo sapi
36	174	13.3	213927	2	AC119391 Rattus no
37	174	13.3	216864	2	AC111592 Rattus no
38	162	12.4	175371	2	AC092822 Homo sapi
39	162	12.4	203003	2	AC087313 Homo sapi
40	145.8	11.2	222	6	AX310312 Sequence
41	145.6	11.2	185388	10	AC122483 Mus muscu
42	137.2	10.5	201194	2	AC139119 Didelphis
43	137.2	10.5	246196	2	AC105641 Rattus no
44	95.8	7.3	10099	1	AE011088 Methanosa
45	83	6.4	11938	1	AE011338 Leptospir

ALIGNMENTS

RESULT 1	BD270234	1305 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	BD270234	Human glycosylation enzymes.			
DEFINITION	BD270234	Human glycosylation enzymes.			
ACCESSION	BD270234.1	GI:33080002			
VERSION	BD270234.1	GI:33080002			
KEYWORDS	JP 2002537796-A/1.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1305)				
AUTHORS	Coleman,T.A. and Betenbaugh,M.J.				
TITLE	Human glycosylation enzymes				
JOURNAL	Patent: JP 2002537796-A 1 12-NOV-2002;				


```

FEATURES
source
Location/Qualifiers
1..1376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
19..1323
/EC_number="2.7.7.43"
/codon_start=1
/product="CMP-N-acetylneuraminic acid synthase"
/protein_id="AA076203.1"

```

```

ORIGIN
Query Match      98.1%; Score 1280.4; DB 9; Length 1305;
Best Local Similarity 99.8%; Pred. No. 4.5e-296;
Matches 1304; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

1  ATGGACTCGTGGAGAAAGGGGCGCGACCTCCGCTCTCCAAACCCGCGGGGGACCGTCC 60
1  ATGGACTCGTGGAGAAAGGGGCGCGACCTCCGCTCTCCAAACCCGCGGGGGACCGTCC 60

61  CGGGCCCGCGCCCGAAAGTGCAGCGCAACTCTCTCGCGCGCGCCAGCGCGAGTGTGGAG 120
61  CGGGCCCGCGCCCGAAAGTGCAGCGCAACTCTCTCGCGCGCGCCAGCGCGAGTGTGGAG 120

121  AAGCCCCCGCACCTGCGACGCCCTAATCTTGGCCCGCGGAGGCGACGAAGGCATCCCCCCTG 180
121  AAGCCCCCGCACCTGCGACGCCCTAATCTTGGCCCGCGGAGGCGAGGAAGGCATCCCCCCTG 180

181  AAGAACAATTAAACACCTGCGGGGGTCCGCTCAATTGCTGGTCTCTGCTGCGGGCCCTG 240
181  AAGAACAATTAAACACCTGCGGGGGTCCGCTCAATTGCTGGTCTCTGCTGCGGGCCCTG 240

241  GATTACGGGGCCCTTCAGAGTGTATGGTTTCACACAGACCATCATGAAATGCGAATGTG 300
241  GATTACGGGGCCCTTCAGAGTGTATGGTTTCACACAGACCATCATGAAATGCGAATGTG 300

301  GCGAACAATTTGGTGCACAAGTTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT 360
301  GCGAACAATTTGGTGCACAAGTTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT 360

361  AACTCACTAGATGCCATCATAGAAATTTCTTAACTATATAAATAGGKTGCATTGTAGGA 420
361  AACTCACTAGATGCCATCATAGAAATTTCTTAACTATATAAATAGGKTGCATTGTAGGA 420

421  AATATTCAGACTACTCTCTCATGTTTACATCTCTACTGATCTTCGAAAGTTGCGAAGATG 480
421  AATATTCAGACTACTCTCTCATGTTTACATCTCTACTGATCTTCGAAAGTTGCGAAGATG 480

481  ATTTCGAGAGAGGATATGATTTCTGTTTCTCTGTTCTGAGACGCCCATCAGTTTCGATGG 540
481  ATTTCGAGAGAGGATATGATTTCTGTTTCTCTGTTCTGAGACGCCCATCAGTTTCGATGG 540

541  AGTGAATTCAGAAAGGAGTTCGTGAAGTGAACGAACTCTCGAATTTAAATCCAGCTAAA 600
541  AGTGAATTCAGAAAGGAGTTCGTGAAGTGAACGAACTCTCGAATTTAAATCCAGCTAAA 600

601  CGGCCCTGTCGACAAAGCTGGATGGAGAAATATATGAAATGCTCATTTTATTTTGCT 660
601  CGGCCCTGTCGACAAAGCTGGGATGGAGAAATATATGAAATGCTCATTTTATTTTGCT 660

661  AAAAGACATTTGATAGAGATGGGTACTCTGCAGGTGG-AAAATGGCATCTACGAAATG 719
661  AAAAGACATTTGATAGAGATGGG---ACTTCAGGGTGAAGAA-GGCMTACTCTAGAAATG 720

720  CGAGCTGGAAACATAGTCTGGATATAGATGTGGATATTGATTGGCCCTATTGCAGAGCAAG 779

```

Thu May 13 11:53:01 2004

/db_xref="GI:8515843"
/translation="MDSVEKGAATSVSNPRGRPRPPKLRNRRGGQGRVKEPPH
LAALLARGSGKIPKNIKHLAGVPLIGWLRALDGAPOSVWSTHDEIENVAK
QFGAQRHRSRVSSTSDAIIEFLNVEHNDIVGNIQATSPCLPTDIDIKVAM
IKREGYDSVSVRRHQRWSEKQGVREVEPLNPAKPRRODWDGELYENGSEFY
FAKHLEMGYLOGGMAYEYENRABHSVDIDIDWPIAEORVLYAGYVGEKLEIK
LIVNIDGLTNGHLYVSGDKEIISYVDKDAIGLSLLKSGISVRLISERACSKTL
SSLKLDCKMEVSVSKLAVDEWRKEMGLCKWEKVEYLVGNEVSEDECCRRKVGUGAPAD
ACSTAQKAVGYICKNGGGRGAIREFAEHICLMEKVNNSQK"

ORIGIN

Query Match 97.5%; Score 1272.4; DB 3; Length 1376;
Best Local Similarity 99.2%; Pred. No. 3.9e-294;
Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;
QY 1 ATGGACTCGGTGAGAGGGGGCCGACCTCGTCTCCAAACCGCGGGGGGACCGTCC 60
DB 19 ATGGACTCGGTGAGAGGGGGCCGACCTCGTCTCCAAACCGCGGGGGGACCGTCC 78
QY 61 CGGGGCGGGCCGCGGAGCTGCAGCGCAACTTCGCGGGGGCCGAGGCGGAGTGGAG 120
DB 73 CGGGGCGGGCCGCGGAGCTGCAGCGCAACTTCGCGGGGGCCGAGGCGGAGTGGAG 138
QY 121 AAGCCCGGCACTGGCAGCCCTAATCTGGCCCGGGGAGGCGAGCAAGCATCCCGCTG 180
DB 139 AAGCCCGGCACTGGCAGCCCTAATCTGGCCCGGGGAGGCGAGCAAGCATCCCGCTG 198
QY 181 AAGAACATTAAGCACCTGGCGGGGGTCCCGCTCATTTGGCTGGTCTGGTGGCGGCTG 240
DB 199 AAGAACATTAAGCACCTGGCGGGGGTCCCGCTCATTTGGCTGGTCTGGTGGCGGCTG 258
QY 241 GATTCAGGGGCGCTCCAGAGTGATGGGTTTCGACGACCATGATGAAATGAGAAATGTC 300
DB 259 GATTCAGGGGCGCTCCAGAGTGATGGGTTTCGACGACCATGATGAAATGAGAAATGTC 318
QY 301 SCCAACAATTTGGTGACAGAGTTCATCGAGAGTTCTGAAGTTTCAAAGAAGCAGCTCT 360
DB 319 SCCAACAATTTGGTGACAGAGTTCATCGAGAGTTCTGAAGTTTCAAAGAAGCAGCTCT 378
QY 361 ACCTCAGTAGATGCATCATAGAAATTTCTTAATTAATATGAGAGTGCATGTAGGA 420
DB 379 ACCTCAGTAGATGCATCATAGAAATTTCTTAATTAATGAGAGTGCATGTAGGA 438
QY 421 AATATTCAGCTACTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCGAAATG 480
DB 439 AATATTCAGCTACTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCGAAATG 498
QY 481 ATTTCGAGAGAGGATATGATCTGKTTCTCTGTTGAGACCCCATCAGTTTCGATGG 540
DB 499 ATTTCGAGAGAGGATATGATCTGTTTCTCTGTTGAGACCCCATCAGTTTCGATGG 558
QY 541 AATGAAATTCAGAAAGGAGTTCGTGAAGTGCAGGAGTCCGAACTCTGAAATTAATCCAGCTAAA 600
DB 559 AATGAAATTCAGAAAGGAGTTCGTGAAGTGCAGGAGTCCGAACTCTGAAATTAATCCAGCTAAA 618
QY 601 CGGCTCTGCGAAGAGTGGATGAGAAATTAATGAAATGCTCAATTTATTTTCT 660
DB 619 CGGCTCTGCGAAGAGTGGATGAGAAATTAATGAAATGCTCAATTTATTTTCT 678
QY 661 AAAAGACATTTGATAGAGATGGTTTACTTCAGGGTGG-AAAATGGCATACTACGAAATG 719
DB 679 AAAAGACATTTGATAGAGATGGTTTACTTCAGGGTGGAAATAAGGCATACGAAATG 738
QY 720 CGAGCTGGAACATAGCTGTGGATATAGATGTGGATATTGATGGGCTATTGACAGCAAG 779
DB 739 CGAGCT-GAACATAGTGGATATAGATGTGGATATTGATGGGCTATTGACAGCAAG 797
QY 780 AGTATAAGATATGCTATTGTCAGAGAGAGCTTAAGGAAATAAAGCTTTGTTG 839
DB 798 AGTATAAGATATGCTATTGTCAGAGAGAGCTTAAGGAAATAAAGCTTTGTTG 857
QY 840 CAATATTCATGCTCTCCACCAAGTGGCCACATTTATGTATCAGGAGACCAAAAGAAAT 899

DB 858 CAATATTCATGCTCTCCACCAAGTGGCCACATTTATGTATCAGGAGACCAAAAGAAAT 917
QY 900 AATATCTTATGATGATAAAGATGCTATTTGGGATAAGTTTATTAAGAAAAGTGGTATTGA 959
DB 918 AATATCTTATGATGATAAAGATGCTATTTGGGATAAGTTTATTAAGAAAAGTGGTATTGA 977
QY 960 GGTGAGGCTTAATCTCAGAAAGGGCCCTGTTCAGAGCAGACGCTGTCTCTTTTAAATCGGA 1019
DB 978 GGTGAGGCTTAATCTCAGAAAGGGCCCTGTTCAGAGCAGACGCTGTCTCTTTTAAATCGGA 1037
QY 1020 TTGCAAAATGGAAGTCAGTGTTCAGACAGCTAGCAGTGTGTAGATGAATGGAGAAAGA 1079
DB 1038 TTGCAAAATGGAAGTCAGTGTTCAGACAGCTAGCAGTGTGTAGATGAATGGAGAAAGA 1097
QY 1080 AATGGCCCTGTCTGAAAGAGAGTGGCATATCTTGAAATGAAAGTGTCTGATGAAAGATG 1139
DB 1098 AATGGCCCTGTCTGAAAGAGAGTGGCATATCTTGAAATGAAAGTGTCTGATGAAAGATG 1157
QY 1140 CTTGAAAGAGAGTGGCCCTAAGTGGCCCTCTGCTGATGCTGTCTTCTTACGCCAGAGGC 1199
DB 1158 CTTGAAAGAGAGTGGCCCTAAGTGGCCCTCTGCTGATGCTGTCTTCTTACGCCAGAGGC 1217
QY 1200 TGTGATACATTTGCAATATGATGTCGCGCCGCTGTCATCCGAGAAATTCGAGAGCA 1259
DB 1218 TGTGATACATTTGCAATATGATGTCGCGCCGCTGTCATCCGAGAAATTCGAGAGCA 1277
QY 1260 CATTTGCTACTAATGAAAGAGTAAATATTCATGCCAAAATAG 1305
DB 1278 CATTTGCTACTAATGAAAGAGTAAATATTCATGCCAAAATAG 1323

RESULT 4

HSMB04286 1766 bp mRNA linear PRI 10-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKFp666I142 (from clone DKFp666I142).
DEFINITION AL832975
ACCESSION AL832975.1 GI:21733563

KEYWORDS

Source Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1766)
AUTHORS Ansoyge W., Winkner U., Mewes H.W., Weil B. and Wiemann S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1. D-85764 Neuberberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.

FEATURES

This clone (DKFp666I142) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/
Location/Qualifiers
1..1766
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFp666I142"
/db_xref="taxon:9606"
/clone="DKFp666I142"
/tissue type="stomach"
/clone_lib="666 (synonym: hsto2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev stage="adult"
1714..1719
1731

ORIGIN

polyA signal
polyA_site
Query Match 97.5%; Score 1272.4; DB 9; Length 1766;
Best Local Similarity 99.2%; Pred. No. 3.9e-294;

Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY 1 ATGGACTCGGTGGAGAGGGGGCGCCACCTCCCGTCTCAACCCCGGGGGGCGACCGTCC 60
 DB 79 ATGGACTCGGTGGAGAGGGGGCGCCACCTCCCGTCTCAACCCCGGGGGGCGACCGTCC 138
 QY 61 CGGGCCCGCGCGCGAAGCTCGAGCGCAACTCTCGCGGGCGGCAAGCGCGAGGTGGAG 120
 DB 139 CGGGCCCGCGCGCGAAGCTCGAGCGCAACTCTCGCGGGCGGCAAGCGCGAGGTGGAG 198
 QY 121 AAGCCCCCGACCTGGAGACCTTAATCTGGCGGGGAGGAGCAAGGCAATCCCGCTG 180
 DB 199 AAGCCCCCGACCTGGAGACCTTAATCTGGCGGGGAGGAGCAAGGCAATCCCGCTG 258
 QY 181 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCAITGGCTGGGTCTCTGGTGGTCTCTGGTGGGCGCTG 318
 DB 259 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCAITGGCTGGGTCTCTGGTGGTCTCTGGTGGGCGCTG 318
 QY 241 GATTTCAGGGGCTTCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 300
 DB 319 GATTTCAGGGGCTTCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 378
 QY 301 GCAAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 360
 DB 379 GCAAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 438
 QY 361 NCTCACTAGATGCCATCATAGATTTCTTAATTAATAATGAAGGKTGACATTTGAGGA 420
 DB 439 NCTCACTAGATGCCATCATAGATTTCTTAATTAATAATGAAGGKTGACATTTGAGGA 498
 QY 421 AATATTCAGACTACTTCTCATGTGTTACATCTACTGATCTTCAAAAGTTGAGAAATG 480
 DB 499 AATATTCAGACTACTTCTCATGTGTTACATCTACTGATCTTCAAAAGTTGAGAAATG 558
 QY 481 ATTGAGAAAGAGGATATGATTTCTGTTTCTGTTGAGGCGCATCAGTTTCATG 540
 DB 559 ATTGAGAAAGAGGATATGATTTCTGTTTCTGTTGAGGCGCATCAGTTTCATG 618
 QY 541 ACTGAAATTCAGAAAGAGTTCGTGAAGTACCGAAGCTCTGTAATTTAAATCCAGCTAA 600
 DB 619 ACTGAAATTCAGAAAGAGTTCGTGAAGTACCGAAGCTCTGTAATTTAAATCCAGCTAA 678
 QY 601 CGGCTCTCGTGCAGACACTGGATGAGAAATATATGAATGGCTCATTTTATTTGCT 660
 DB 679 CGGCTCTCGTGCAGACACTGGATGAGAAATATATGAATGGCTCATTTTATTTGCT 738
 QY 661 AAAAGACATTTGATAGAGATGGTTACTTTCAGGGTGG-AAAATGCGATACACGAAATG 719
 DB 739 AAAAGACATTTGATAGAGATGGTTACTTTCAGGGTGGAAATATGCGATACACGAAATG 798
 QY 720 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATTTGATTTGCGCTATTTCAGAGCAAG 779
 DB 799 CGAGCT-GAACATAGTGTGGATATAGATGTGGATATTTGATTTGCGCTATTTCAGAGCAAG 857
 QY 780 AGTATTAAGATATGGCTATTTTGGGAAAGAGCTTACGAAATTAACCTTTGTTG 839
 DB 858 AGTATTAAGATATGGCTATTTTGGGAAAGAGCTTACGAAATTAACCTTTGTTG 917
 QY 840 CAATATGATGATGCTCAACATGCGCACTTTATGATATCAGGAGCAACAAAGAAAT 899
 DB 918 CAATATGATGATGCTCAACATGCGCACTTTATGATATCAGGAGCAACAAAGAAAT 977
 QY 900 AATCATCTTATGATTAAGATGCTATTGGGATAAGTTTATTAAGAAAGAGTGGTATGA 959
 DB 978 AATATCTTATGATTAAGATGCTATTGGGATAAGTTTATTAAGAAAGAGTGGTATGA 1037
 QY 960 GGTGAGGCTAATCTCAGAAAGGCGCTGTTTCAAGCAGACGCTGCTCTTTTAAACCTGGA 1019
 DB 1038 GGTGAGGCTAATCTCAGAAAGGCGCTGTTTCAAGCAGACGCTGCTCTTTTAAACCTGGA 1097
 QY 1020 TTGCAAAATGGAAGTCAAGTATCAGCAAGCTAGCAGTTGTAGTCAATGGAAGAA 1079
 DB 1098 TTGCAAAATGGAAGTCAAGTATCAGCAAGCTAGCAGTTGTAGTCAATGGAAGAA 1157

RESULT 5
 AR338869
 LOCUS
 REFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 1 (bases 1 to 1770)
 AUTHORS
 Tang, Y.T., Zhou, P. and Drmanac, R.T.
 TITLE
 Nucleic acids and polypeptides
 JOURNAL
 Patent: US 659662-A 360 27-MAY-2003;
 FEATURES
 1..1770
 Location/Qualifiers
 source
 /organism="unknown"
 /mol_type="genomic DNA"

Query Match 97.5%; Score 1272.4; DB 6; Length 1770;
 Best Local Similarity 99.2%; Pred. No. 4e-294;
 Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY 1 ATGGACTCGGTGGAGAGGGGGCGCCACCTCCCGTCTCAACCCCGGGGGGCGACCGTCC 60
 DB 105 ATGGACTCGGTGGAGAGGGGGCGCCACCTCCCGTCTCAACCCCGGGGGGCGACCGTCC 164
 QY 61 CGGGCCCGCGCGCGAAGCTTCAGAGCGCAACTCTCGCGGGCGGCGAGGCTGGAG 120
 DB 165 CGGGCCCGCGCGCGAAGCTTCAGAGCGCAACTCTCGCGGGCGGCGAGGCTGGAG 224
 QY 121 AAGCCCCCGACCTGGAGACCTTAATCTGGCGGGGAGGAGCAAGGCAATCCCGCTG 180
 DB 225 AAGCCCCCGACCTGGAGACCTTAATCTGGCGGGGAGGAGCAAGGCAATCCCGCTG 284
 QY 181 AAGAACATTAAGCACTGGCGGGGTCCTCGCTCAITGGCTGGGTCTCGTGGCGCTG 240
 DB 285 AAGAACATTAAGCACTGGCGGGGTCCTCGCTCAITGGCTGGGTCTCGTGGCGCTG 344
 QY 241 GATTTCAGGGGCTTCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 300
 DB 345 GATTTCAGGGGCTTCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAAATG 404
 QY 301 GCAAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 360
 DB 405 GCAAAACAAATTTGGTGCACAGTTCATCGAAGATTTCTGAAGTTTCAAAAGACAGCTCT 464
 QY 361 ACCTCACTAGATGCCATCATAGATTTCTTAATTAATAATGAGGKTGACATTTGAGGA 420
 DB 465 ACCTCACTAGATGCCATCATAGATTTCTTAATTAATAATGAGGKTGACATTTGAGGA 524
 QY 421 AATATTCAGACTACTTCTCATGTGTTACATCTACTGATCTTCAAAAGTTGAGAAATG 480
 DB 525 AATATTCAGACTACTTCTCATGTGTTACATCTACTGATCTTCAAAAGTTGAGAAATG 584

JOURNAL		Patent: EP 1074617-A 17324 07-FEB-2001;	
FEATURES		Research Association for Biotechnology (JP)	
source		Location/Qualifiers	
		1..1694	
		/organism="Homo sapiens"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:9606"	
CDS		41..1345	
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="CAE91573.1"	
		/db_xref="GI:40037279"	
		/translation="MDSVEKGAATSVNPRPSRGRPPKPKLRNRRGGQGRGVEKPKPH	
		LAALILARGSGKIPKIKHLAGVPLIGWLRALDSDGAPQSVWVSTHDELENVAK	
		OFGAQRHSSEKSDSLDAIIEFLNYHNEVDIVGNQATSPCLHDDLDQVASM	
		IEEGYDVSFVRHQRWSEVTEPLNPKRPRRQWDGDELYENGYSY	
		FAKHLLIEMGYLQGRMAYEMEAHSVDIDWDPIARQRLVRYGPGKEKLKIK	
		LIVCNIDGCLTNGHIIVSGDKELISYDVKDAIGISLLKSGIEVLISERACSKOTL	
		SSLKLDCKMEVSVSDKLAIVDWEKMGELCKEVAIILGNESVDEECLKRVLSGAPAD	
		ACSTAQKAVGYICKNGSGGAIRFPAHICLIMEKVNNSQK"	
ORIGIN			
Query Match		97.4%; Score 1270.8; DB 6; Length 1694;	
Best Local Similarity		99.1%; Pred. No. 9.5e-294;	
Matches 1294;		Conservative 4; Mismatches 6; Indels 2; Gaps 2;	
QY	1	ATGACTCGGTGGAGAGGGGGCGCCACCTCCCTCTCCAACTCCGCGGGGGCGACCGTCC	60
DB	41	ATGACTCGGTGGAGAGGGGGCGCCACCTCCCTCTCCAACTCCGCGGGGGCGACCGTCC	100
QY	61	CGGGCGGGCGCGCGAGCTGCGAGGMACTCTCGGGCGGGCGGGCGGGGCTGGAG	120
DB	101	CGGGCGGGCGCGCGAGCTGCGAGGMACTCTCGGGCGGGCGGGCGGGGCTGGAG	160
QY	121	AAGCCCGCGACCTGGCAGCCCTTAATCTGCGCGGGGAGCGCAGCAAGGCAATCCCTG	180
DB	161	AAGCCCGCGACCTGGCAGCCCTTAATCTGCGCGGGGAGCGCAGCAAGGCAATCCCTG	220
QY	181	AAGAACATTAAAGCACTTCGCGGGGGTCCGCTCAATTCGCTGGTCTCTGCGGGCGCTG	240
DB	221	AAGAACATTAAAGCACTTCGCGGGGGTCCGCTCAATTCGCTGGTCTCTGCGGGCGCTG	280
QY	241	GATTGAGGGGCTTCGAGAGTGTATGGTTTCGAGACACCATGATGAATTCAGATG	300
DB	281	GATTGAGGGGCTTCGAGAGTGTATGGTTTCGAGACACCATGATGAATTCAGATG	340
QY	301	GCCAAACAATTTGGTGCACAGTTCATCGAAGAAGTCTGAAAGTTTCAAAAGACAGCTCT	360
DB	341	GCCAAACAATTTGGTGCACAGTTCATCGAAGAAGTCTGAAAGTTTCAAAAGACAGCTCT	400
QY	361	ACCTCCTAGATGCCATCATAGAAATTTCTTAATTAATTAATTAATTAATTAATTAATTA	420
DB	401	ACCTCCTAGATGCCATCATAGAAATTTCTTAATTAATTAATTAATTAATTAATTAATTA	460
QY	421	ATATTCAGCTACCTTCYCATGTTTACATCTCTACATCTCTCAAAAGTTCGAGAAAG	480
DB	461	ATATTCAGCTACCTTCYCATGTTTACATCTCTACATCTCTCAAAAGTTCGAGAAAG	520
QY	481	ATTCGAGAGAGAGATGATTCCTGCTGTTTGTGAGACGCGCATCAGTTTCGATGG	540
DB	521	ATTCGAGAGAGAGATGATTCCTGCTGTTTGTGAGACGCGCATCAGTTTCGATGG	580
QY	541	AGTGAATTCAGAAAGAGAGTTCGTGAAGTGCAGCAACCTCTCAATTAATTAATTAATTA	600
DB	581	AGTGAATTCAGAAAGAGAGTTCGTGAAGTGCAGCAACCTCTCAATTAATTAATTAATTA	640
QY	601	CGGCTTCGTCGACAGCTGGGATGAGATTTATATGAATGGCTCATTTATTTTGTCT	660
DB	641	CGGCTTCGTCGACAGCTGGGATGAGATTTATATGAATGGCTCATTTATTTTGTCT	700
QY	661	AAAGACATTTCTAGAGTGGTTCCTGAGGGTGG-AAAATGCGATCTACTACGAAATG	719
DB	701	AAAGACATTTCTAGAGTGGTTCCTGAGGGTGG-AAAATGCGATCTACTACGAAATG	760

	FT	CDS	(41)...(1342)
720	CGAGCTGGACATGATGTGGATATAGATGTGGTATTTGATTCGCCTATTGCCAGCAAGA	779	
761	CGAGCT-GAACATAGTGTGGATATAGATGTGGTATTTGATTCGCCTATTGCCAGCAAGA	819	
780	AGTATTAAGATATGGCTATTTTGCAAGAGAGAGCTTAAGGAATAAACAATTTGGTTTG	839	
820	AGTATTAAGATATGGCTATTTTGCAAGAGAGAGCTTAAGGAATAAACAATTTGGTTTG	879	
840	CAATATTAAGATATGGCTATTTTGCAAGAGAGAGCTTAAGGAATAAACAATTTGGTTTG	899	
880	CAATATTAAGATATGGCTATTTTGCAAGAGAGAGCTTAAGGAATAAACAATTTGGTTTG	939	
900	AATATCTTATGATTAAGATATGGCTATTTTGCAAGAGAGAGCTTAAGGAATAAACAATTTGGTTTG	959	
940	AATATCTTATGATTAAGATATGGCTATTTTGCAAGAGAGAGCTTAAGGAATAAACAATTTGGTTTG	999	
960	GATGAGCTAACTCAGAAAAGGCTCTTCAAAGCAGACGCTGCTCTTTTAAACTGGA	1019	
1000	GATGAGCTAACTCAGAAAAGGCTCTTCAAAGCAGACGCTGCTCTTTTAAACTGGA	1059	
1020	TTCGAAATGGAAGCTAGTGTATCAGACAAAGCTAGAGTGTAGTAGAATGGAGAAAGA	1079	
1060	TTCGAAATGGAAGCTAGTGTATCAGACAAAGCTAGAGTGTAGTAGAATGGAGAAAGA	1119	
1080	AATGGGCTGTGCTGGAAAGAGTGGCATCTTTGGAAATGAAGTGTCTGATGAAGAGTG	1139	
1120	AATGGGCTGTGCTGGAAAGAGTGGCATCTTTGGAAATGAAGTGTCTGATGAAGAGTG	1179	
1140	CTTCAAGAGAGTGGCCCTAAGTGGGCTCTCTGCTGATGCTGCTCTACGCCAGAGGC	1199	
1180	CTTCAAGAGAGTGGCCCTAAGTGGGCTCTCTGCTGATGCTGCTCTACGCCAGAGGC	1239	
1200	TGTTGGATACATTTGCAAAATGTAATGTTGGGCTGGTGGCTCCATCCGAGAAATTTGCAGAC	1259	
1240	TGTTGGATACATTTGCAAAATGTAATGTTGGGCTGGTGGCTCCATCCGAGAAATTTGCAGAC	1299	
1260	CATTGGCTACTAATGGAAGAGTAAATTAATCATGCCCAAAAAATAG	1305	
1300	CATTGGCTACTAATGGAAGAGTAAATTAATCATGCCCAAAAAATAG	1345	
RESULT 7	BD159714	1694 bp	DNA linear PAT 17-JAN-2003
LOCUS	Primer for synthesizing full-length cDNA and use thereof.		
DEFINITION	BD159714		
ACCESSION	BD159714.1	GI:27865472	
VERSION	JP 2002191363-A/14557.		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,		
JOURNAL	Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.		
COMMENT	Patent: JP 2002191363-A 14557 09-JUL-2002;		
	HELIX RESEARCH INSTITUTE		
	OS Homo sapiens (human)		
	PN JP 2002191363-A/14557		
	PD 09-JUL-2002		
	PF 28-JUL-2000 JP 2000280990		
	PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU		
	PI SAITO,		
	PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,		
	PI KEIICHI NAGAI, TETSUO OTSUKI		
	PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/13, C12N1/21, C12N5/ PC		
	10,		
	PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC		
	Primer for synthesizing full-length cDNA and use thereof FH Key		
	Location/Qualifiers		

```

FEATURES
  source
    Location/Qualifiers
      1. .1694
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="NT2RP2003643"
        /cell_line="NT2"
        /cell_type="teratocarcinoma"
        /clone_id="NT2RP2"
        /notes="cloning vector: pME18SFL3-mRNA from NT2 neuronal
        precursor cells after 2-weeks retinoic acid (RA)"

```

CDS	induction."	Query Match	Score	DB	Length
41..1345	/notes="unnamed protein product"	97.4%	1270.8	DB 9	1694
/codon_start=1	/protein_id="BAB14311.1"	Best Local Similarity	99.1%	Prod. No. 9.5e-294	
/db_xref="GI:10434599"		Matches 1294	Conservative	4	Mismatches 6
/translation="MDSVRKGAATSVSNPRGRSPRPKLQNSRGSGGQGVKEKPPH LMAILLARGGSKGPIPLKNIKHLAGVPLIGWLVFAALDSGAFQSVWATSPCLTHDELQKVAEM QGAQVHRSSVESKDSSTSLDAILIEFLYNHNEVDIVGNTOATSPCLTHDELQKVAEM IREQGVSVFVSVRRHQSFTKQGVREVTEVIN.NPAKPRPRQMDWGLYENGSFY FAKRLHENGVLQGGHIVYMEAEHSVDIDVDIMPIAEQRVLRVYGFQKELKEIKL LLVNCNIDGLTNGHIVYVSGDOKIISVDVNDKIGISLLKSGIEVFLISERRASGKQTL SSLLCKKEVSVSDKLVAVDEWRKMGILCKEYAVLGNESVDEECLIKRVGLSGAPAD ACSTAQKAVGTYICKNGRGRAIRFPAEHLCLLMEKYNNSQK"					
ORIGIN					
Query Match	97.4%	Score 1270.8	DB 9	Length 1694	
Best Local Similarity	99.1%	Prod. No. 9.5e-294			
Matches 1294	Conservative	4	Mismatches 6	Indels 2	Gaps 2
QY	1	ATGACTCGTGGAGAGGGGGCGCACCTCCGTCTCCACCCGCGGGGGCGAGCGTCC	60		
DB	41	ATGACTCGTGGAGAGGGGGCGCGCACCTCCGTCTCCACCCGCGGGGGCGAGCGTCC	100		
QY	51	CGGGCGGGCGCGCGAGCTCGAGCGCAACTCTCGCGCGGCGCCAGGGCGAGGTTGGAG	120		
DB	101	CGGGCGGGCGCGCGAGCTCGAGCGCAACTCTCGCGCGGCGCCAGGTCGAGGTTGGAG	160		
QY	121	AAGCCCGCGACCTCGAGCGCTTAATTTCTGCGCGGGAGCGAGAGGGCTCCCCCTG	180		
DB	161	AAGCCCGCGACCTCGAGCGCTTAATTTCTGCGCGGGAGCGAGAGGGCTCCCCCTG	220		
QY	181	AAGAACATTAAAGCACTGGCGGGGGTCCGCTCATTTGGCTGGGTCTCGGTGGCGCCCTG	240		
DB	221	AAGAACATTAAAGCACTGGCGGGGGTCCGCTCATTTGGCTGGGTCTCGGTGGCGCCCTG	280		
QY	241	GATTCAGGGCGCTTCAGAGTGTTGGTTTCGACAGACCATGATGAATTCAGAACTGG	300		
DB	281	GATTCAGGGCGCTTCAGAGTGTTGGTTTCGACAGACCATGATGAATTCAGAACTGG	340		
QY	301	GCCAAACAATTTGGTGCAACAAGTCTCATGAAAGTCTGAGTTTCAAAGACAGCTCT	360		
DB	341	GCCAAACAATTTGGTGCAACAAGTCTCATGAAAGTCTGAGTTTCAAAGACAGCTCT	400		
QY	361	ACCTCACTAGATGCCATCAGAAATTCCTTAATTAATTAATGAGGKTGCACATTTGAGA	420		
DB	401	ACCTCACTAGATGCCATCAGAAATTCCTTAATTAATTAATGAGGKTGCACATTTGAGA	460		
QY	421	ATATTTCAAGTACTCTTCATGTTTACATCTCTATGATCTTCAAAGAGTTCGACAAATG	480		
DB	461	ATATTTCAAGTACTCTTCATGTTTACATCTCTATGATCTTCAAAGAGTTCGACAAATG	520		
QY	481	ATTTCGAGAGAGGATATGATTTCTGTTTCTGTTGAGACGGCATCAGTTTCGATGG	540		
DB	521	ATTTCGAGAGAGGATATGATTTCTGTTTCTGTTTGTGAGACGGCATCAGTTTCGATGG	580		
QY	541	AGTGAATTCAGAAAGGAGTTCGTAAGTGCACCGACCTCTGAATTTAAATCCAGCTAAA	600		
DB	581	AGTGAATTCAGAAAGGAGTTCGTAAGTGCACCGACCTCTGAATTTAAATCCAGCTAAA	640		
QY	601	CGGCTCTGTCGACAGACTGGGATGAGAAATTAATGAAATTTGGCTCATTTATTTGCT	660		
DB	641	CGGCTCTGTCGACAGACTGGGATGAGAAATTAATGAAATTTGGCTCATTTATTTGCT	700		
QY	661	AAAAGACATTTGATAGAGATGGGTTACTTTGCGGGTGG-AAAATGCACTACTACGAAATG	719		
DB	701	AAAAGACATTTGATAGAGATGGGTTACTTTGCGGGTGGAAATGATGAAATG	760		
QY	720	CGAGCTTGGACATAGTGTGGATATAGATGTGGATTTGATTTGGCTTATTCGAGCAAAG	779		
DB	761	CGAGCTTGGACATAGTGTGGATATAGATGTGGATTTGATTTGGCTTATTCGAGCAAAG	819		
QY	780	AGTATTAAGATATGGCTATTTTGGCAAGNGAGCTTAAGGAAATAAAACTTTTGGTTTG	839		

Position	Sequence	Position	Sequence	Position	Sequence
360	GGT GAGGCTAAATCTCAGAAAGGGCCCTGTTCAAAGCAGACGCCTGCTCTTTTAAAACCTGGA	1019			
1060	GGT GAGGCTAAATCTCAGAAAGGGCCCTGTTCAAAGCAGACGCCTGCTCTTTAAAACCTGGA	1119			
1020	TTGCAAAATGGAAGTCAGTGTATCAGACAAGCTAGCAGTTGTAGATGAATGGAGAAAAGA	1079			
1120	TTGCAAAATGGAAGTCAGTGTATCAGACAAGCTAGCAGTTGTAGATGAATGGAGAAAAGA	1179			
1080	AATGGCCCTGTGCTGCGAAGAAAGTGGCAATATCTTGAANAAGAGTGTCTGATGAAGAGTG	1139			
1180	AATGGCCCTGTGCTGCGAAGAAAGTGGCAATATCTTGAANAAGAGTGTCTGATGAAGAGTG	1239			
1140	CTTGAAGAGAGTGGGCTTAAGTGGCGCTCTGCTGATGCGCTGTCTCTAGCCCCAGAGGC	1199			
1240	CTTGAAGAGAGTGGGCTTAAGTGGCGCTCTGCTGATGCGCTGTCTCTAGCCCCAGAGGC	1299			
1200	TGTTGGATACATTTGCAAAATGTAATGGTGGCCCTGGTGGCCATCCGAGAAATTTGCAGACA	1259			
1300	TGTTGGATACATTTGCAAAATGTAATGGTGGCCCTGGTGGCCATCCGAGAAATTTGCAGACA	1359			
1260	CATTTGCCCTACTAAATGGAAAAAGTTAATTAATTCATGCCAAAAATAG	1305			
1360	CATTTGCCCTACTAAATGGAAAAAGTTAATTAATTCATGCCAAAAATAG	1405			

```

RESULT 10
AXC40083
LOCUS          1772 bp          linear          PAT 18-NOV-2000
DEFINITION     Sequence 6 from Patent WO0063351.
ACCESSION      AXC40083
VERSION        AXC40083.1  GI:11230045
KEYWORDS
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1. Lal, P., Yue, H., Tang, Y., Hillman, J.L., Baughn, M.R. and Yang, J.
                Carbohydrate-modifying enzymes
                Patent: WO 0063351-A 5 26-OCT-2000;
                Incyte Genomics, Inc. (US)
                Location/Qualifiers
                1. 1772
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
                /note="Incyte ID No: C00422CBI"
FEATURES
source

```

Query Match	97.4%	Score 1270.8	DB 6	Length 1772
Best local Similarity	99.1%	Pred. No. 9.6e-294	Indels 2	Gaps 2
Matches 1294	Conservative 4	Mismatches 6		
QY	1	ATGGACTCGGTGGAGAGAGGGGGCGCCACCTCCGGTCTCCAAACCCCGGGGGCGACCGTCC	60	
Db	101	ATGGACTCGGTGGAGAGAGGGGGCGCCACCTCCGGTCTCCAAACCCCGGGGGCGACCGTCC	160	
QY	61	CGGGGCGGGCGCGCGAAGCTCAGCGCAACTCTCGCGCGGCCAGGGCGGAGGTGTGGAG	120	
Db	161	CGGGGCGGGCGCGCGAAGCTCAGCGCAACTCTCGCGCGGCCAGGGCGGAGGTGTGGAG	220	
QY	121	AAGCCCCCGACCTGGCAGCCCTTAATTGGGCCCGGAGGCGAGCAAGGCATCCCCCTG	180	
Db	221	AAGCCCCCGACCTGGCAGCCCTTAATTGGGCCCGGAGGCGAGCAAGGCATCCCCCTG	280	
QY	181	AAGAACATTAAAGACACTGGCGGGGGGTCCGGCTCATGTGGTGGTCTCGTGGCGCCCTG	240	
Db	281	AAGAACATTAAAGACACTGGCGGGGGGTCCGGCTCATGTGGTGGTCTCGTGGCGCCCTG	340	
QY	241	GATTGAGGGGCTTCAGAGTGTATGGGTTTCGACGAGCACCATGATGAAATTCAGAAATGTG	300	
Db	341	GATTGAGGGGCTTCAGAGTGTATGGGTTTCGACGAGCACCATGATGAAATTCAGAAATGTG	400	

[illegible]

RESULT 11
BC063776
LOCUS
DEFINITION

BC063776 1704 bp mRNA linear ROD 16-DEC-2003
MGC:70211 IMAGE:6398450, complete cds.

```

ACCESSION      BC063776
VERSION        BC063776.1  GI:39795344
KEYWORDS       Mus musculus (house mouse)
SOURCE         MGC.
ORGANISM       Mus musculus
REFERENCE
AUTHORS        Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
                Klausner, R.D., Collins, P.S., Wagner, C.M., Schuler, G.D.,
                Altschul, S.F., Zeeberg, K.H., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
                Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
                Ditzchenko, L., Marzina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
                Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.J.,
                Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
                Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
                Abramson, R.D., Mulany, S.J., Bosak, S.A., McEwan, P.J.,
                McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
                Woxley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
                Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
                Pahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
                Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
                Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
                Dickinson, V.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
                Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.B.,
                Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
                Generation and initial analysis of more than 15,000 full-length
                human and mouse cDNA sequences
                Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002);
                12477932
                2 (bases 1 to 1704)
                Direct Submission
                Submitted (03-DEC-2003) National Institutes of Health, Mammalian
                Gene Collection (MGC), Cancer Genomics Office, National Cancer
                Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
                USA
                NIH-MGC Project URL: http://mgc.nci.nih.gov
                Contact: MGC help desk
                Email: cgabbs@mail.nih.gov
                Tissue Procurement: Jeffrey E. Green, M.D.
                CDNA Library Preparation: Life Technologies, Inc.
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Genome Sequence Centre,
                BC Cancer Agency, Vancouver, BC, Canada
                info@bcgsc.bc.ca
                Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
                Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
                Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
                Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
                Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
                Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
                Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Varçy,
                George Yang, Scott Zuyderduyn, Marco Marra.
                Clone distribution: MGC clone distribution information can be found
                through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                Series: IRAC Plate: 131 Row: n Column: 21
                This clone was selected for full length sequencing because it
                passed the following selection criteria: matched mRNA gi: 22238853.
                Location/Qualifiers
                1. .1704
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="MGC:70211 IMAGE:6398450"
                /tissue types="Colon, normal. 5 month old male mouse."
                /clone_lib="NCI CGAP Co24"
                /lab_host="DH10B"
                /note="vector: pCMV-SPORT6"
                17. .1315
                /codon_start=1
                CDS
                ORIGIN
                Query Match 77.3%; Score 1008.8; DB 10; Length 1704;
                Best Local Similarity 87.1%; Pzed. No. 6.5e-231;
                Matches 1138; Conservative 4; Mismatches 156; Indels 8; Gaps 3;
                1 ATGGACTCGGTGGAGAGGGGGCGGCACCTCGTCTCCACCCGCGGGGGGACCGTCC 60
                17 ATGGAGCGCTCGAGAGAGGGGGCGGCACCTCGTCTCCACCCGCGGGGGGACCGTCC 76
                61 CGGGGGCGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
                77 CGGGGGCGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 130
                121 AAGCCCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
                131 AAGCCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 190
                181 AAGAACATTAAAGCAGCGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
                191 AAGAACATCAAGCGCGCTCGGGGGGCTCCGCTCATTTGGCTGGGCTCCGCGCGCGCG 250
                241 GATTGAGGGCGCTCCGAGAGTGATGGGTTTCACACACCATCATGATGAAATGAGAATGTG 300
                251 GATCGGGGGGCTCTCCAGAGTGATGGGTTTCACACACCATCATGATGAAATGAGAATGTG 310
                301 GCCAAACAAATTTGGTGACCAAGTTTCATCGAAGAGTTTCTGAAGTTTCAAAGACAGCTCT 360
                311 GCCAAACAGTTTGGTGACCAAGTTTCATCGAAGAGTTTCTGAAGTTTCAAAGACAGCTCT 370
                361 ACCTCACTAGAGCCATCATAGAAATTTCTTAATATATATATATAGAGGAGGAGTTCGATG 420
                371 ACCTCACTAGAGCCATCATAGAAATTTCTTAATATATATATATAGAGGAGGAGTTCGATG 430
                421 AATATCAAGCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
                431 AATATCAAGCCCATCTCCATGTTTACATCCCACTCCCACTCCCAAGAGTTGCAAGANTG 490
                481 ATTTCGAGAGAGGATATGATTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 540
                491 ATCCGAGAGAGGATATGATTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 550
                541 AGTGAATTCAGAAAGAGTTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAG 600
                551 AGTGAATTCAGAAAGAGTTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAG 610
                601 CGGCTCTGTCAGCAAGAGTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAG 660
                611 CGGCTCTGTCAGCAAGAGTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAGAGTTCGTTGAG 670
                661 AAGAGACATTGATAGAGATGGGTTTACTTTCAGGGTTCGTTGAGAGTTCGTTGAGAGTTC 719

```



```

Db      671  AAAAGCAATTGATAGAGATGGTACTTACAGGGTGGGAAATGCGTATTATGAATG 730
QY      720  CGAGCTGGAACATAGTGTGGATAGATGGATATGATGGCTTATTCAGAGCAAG 779
Db      731  CGAGCT-GAGCAGTGTGGATATGAGTGGATCGAGTCGATGCGGATCGAGCAAG 789
QY      780  AGTATTAGATATGGCTATTTGGCAAGAGAGAGCTTAAGCAATATAAACTTTGGTTG 839
Db      790  AGTCTCAGATTGGCTATTTGGCAAGAGAGAGCTTAAGCAATATAAACTTTGGTTG 849
QY      840  CAATATTGATGATGTCTCCAAATGCGCCCHTTATGATCAGGACACCAAAAGAAAT 899
Db      850  TAATATTGATGGATGCTCTCCAAATGCGCCCAATTTATGATCAGGACACCAAAAGAAAT 909
QY      900  AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959
Db      910  AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 969
QY      960  GGTGAGGCTAATCTCAGAAAGGGCTCTTCAAGCAGACGCTGCTCTTTTAAACTGGA 1019
Db      970  GGTGAGGCTCATCTCAGACGGGCTCTCTCCAGCAGACGCTCTCCGCGCTTAAAGCTGA 1029
QY      1020  TCGCAAAATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1079
Db      1030  CTGTAAACAGAGTCAAGTGTGTCCGATAAGCTGACCGCCACCGTGTGATGATGATG 1089
QY      1080  AATGGGCTGTCTGGAAGAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGATGATG 1139
Db      1090  GATGGGCTGTCTGGAAGAGAGTGGCATATCTTGGCAATGAAGTGTCTGATGATGATG 1149
QY      1140  CTTCAAGAGATGGGCTTAAGTGGCTCTCTGATGATGATGATGATGATGATGATGATG 1199
Db      1150  CTTCAAGAGATGGGCTTAAGTGGCTCTCTGATGATGATGATGATGATGATGATGATG 1209
QY      1200  TGTGGATACATTTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1259
Db      1210  TGTGGGATACATCTGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATG 1269
QY      1260  CATTTGCTTACTAATGGAAGAGTAAATAATCTCATGCGCAAAATAG 1305
Db      1270  CATTTGCTTACTGATAGAAAGTAAATAATCTCATGCGCAAAATAG 1315

RESULT 12
LOCUS   BC031500
DEFINITION Mus musculus cytidine monophospho-N-acetylneuraminic acid
synthetase, mRNA (cdna clone MGC:27736 IMAGE:2647633), complete
cds.
ACCESSION BC031500
VERSION   BC031500.1 GI:2619375
KEYWORDS MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1728)
AUTHORS   Strausberg,R.L., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Klausner,R.T., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Altschul,S.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,E.,
Hopkins,R.F., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Diatcherko,L., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Stapleton,M., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Guaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.C., Blakesley,R.W., Touchman,J.W., Green,E.D.,

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
Contact: MGC help desk
Email: gcapbs-rc@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Guenatne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lottsege, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC plate: 35 Row: b Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22208853.
Location/Qualifiers
1..1728
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:27736 IMAGE:2647633"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP Mam"
/lab_host="DHCE"
/notes="Vector: pCMV-SPORT6"
1..1728
/gene="Cmas"
/notes="synonym: CMP-Neu5Ac"
/db_xref="LocusID:12764"
/db_xref="MGI:1337124"
17..1315
/codon_start=1
/product="cytidine monophospho-N-acetylneuraminic acid
synthetase"
/protein_id="AAH31500.1"
/db_xref="GI:21619376"
/db_xref="LocusID:12764"
/translation="MDALEKGSATSPAPRGSRPGRPKLQSRGAGRGKPPHLLA
ALIVARGSGKIPKRIKRIAGVPLIGVIRALDAGVFSQVSVSTHDIENVAQKF
CAQVHRSSTSKDSSTDAIVEFLYHNEVDVGNIOATSPCLHTDLOKVAEMIR
EGYDSVFSVRHQFWSIEIOKGVRETEPLNAPKPRQDWDGELKNGSFYFA
KRLHEMGLQGVKMYEMRAHSVDIVDIPAEORVIRFGFYGKELKEIKLL
VCNIDGCLTNGHIVYSGDQKEIISYDVDAIGSILKSGIEVRLSEACSKQTLGA
LKDDCKTWSVSDKATVDWEKMLCWKEVLYGNEVSDDECLKRVGLSAVPAADAC
SGAQAAGVYICKSGRGRAIRFAEHLILIEKWNVCQK"
146..781
/notes="CTP transf 3; Region: Cytidyltransferase. This
family consists of two main Cytidyltransferase
activities: 1) 3-deoxy-manno-octulosonate
Cytidyltransferase, EC:2.7.7.38 catalysing the
reaction:- CTP + 3-deoxy-D-manno-octulosonate <=>

```

```

Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Batterfield,Y.S., Krzywinski,M.I., Skalska,J., Smalish,D.E.,
Schnerich,A., Schein,J.E., Jones,S.J. and Marita,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1728)
Strausberg,R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcapbs-rc@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Guenatne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lottsege, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC plate: 35 Row: b Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22208853.
Location/Qualifiers
1..1728
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:27736 IMAGE:2647633"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP Mam"
/lab_host="DHCE"
/notes="Vector: pCMV-SPORT6"
1..1728
/gene="Cmas"
/notes="synonym: CMP-Neu5Ac"
/db_xref="LocusID:12764"
/db_xref="MGI:1337124"
17..1315
/codon_start=1
/product="cytidine monophospho-N-acetylneuraminic acid
synthetase"
/protein_id="AAH31500.1"
/db_xref="GI:21619376"
/db_xref="LocusID:12764"
/translation="MDALEKGSATSPAPRGSRPGRPKLQSRGAGRGKPPHLLA
ALIVARGSGKIPKRIKRIAGVPLIGVIRALDAGVFSQVSVSTHDIENVAQKF
CAQVHRSSTSKDSSTDAIVEFLYHNEVDVGNIOATSPCLHTDLOKVAEMIR
EGYDSVFSVRHQFWSIEIOKGVRETEPLNAPKPRQDWDGELKNGSFYFA
KRLHEMGLQGVKMYEMRAHSVDIVDIPAEORVIRFGFYGKELKEIKLL
VCNIDGCLTNGHIVYSGDQKEIISYDVDAIGSILKSGIEVRLSEACSKQTLGA
LKDDCKTWSVSDKATVDWEKMLCWKEVLYGNEVSDDECLKRVGLSAVPAADAC
SGAQAAGVYICKSGRGRAIRFAEHLILIEKWNVCQK"
146..781
/notes="CTP transf 3; Region: Cytidyltransferase. This
family consists of two main Cytidyltransferase
activities: 1) 3-deoxy-manno-octulosonate
Cytidyltransferase, EC:2.7.7.38 catalysing the
reaction:- CTP + 3-deoxy-D-manno-octulosonate <=>

```

diphosphate + CMP-3-deoxy-D-manno-octulosonate, 2)
acylneuraminate cytidyltransferase EC:2.7.7.43,
catalysing the reaction:- CTP + N-acylneuraminate <=>
diphosphate + CMP-N-acylneuraminate"
/db_xref="CDD:pfam02348"
818..1273
misc_feature
/note="COG1778; Region: Low specificity phosphatase (HAD
superfamily) [General function prediction only]";
/db_xref="CDD:COG1778"

ORIGIN

Query Match 77.3%; Score 1008.8; DB 10; Length 1728;
Best Local Similarity 87.1%; Pred. No. 6.5e-231;
Matches 1138; Conservative 4; Mismatches 156; Indels 8; Gaps 3;

QY 1 ATGACTCGTGAGAGAGGGGGCGGACCTCGCTCCACACCGCGGGGGCGACCGTCC 60
DB 17 ATGACGCGCTGAGAGAGGGGGCGGACCTCGCTCGGCGCGCGCGCGCGCGTCC 76
QY 61 CGGGGGCGGGCGCGGAGAGCTGAGCGCACTCTCGCGCGCGCGCGCGCGCGTGGAG 120
DB 77 CGGGGGCGGGCGCGGAGAGCTGAGCGCA-----GCGGGGGCGCGCGCGCTAGAG 130
QY 121 AAGCCCGCGCACCTGGCAGCCCAATTCCTGGCGGGGGAGGCGAGCAAGGCAATCCCGCTG 180
DB 131 AAGCCCGCGCACCTGGCAGCGCTGGTCTGGCGCGCGCGCGCGCGCAAGGCAATCCCACTG 190
QY 181 AAGAACATTAAGCACCTGGCGGGGGTCCGCTCATTTGGCTGGGTCTCGCGTGGGGCGCTG 240
DB 191 AAGAACATTAAGCACCTGGCGGGGGTCCGCTCATTTGGCTGGGTCTCGCGTGGGGCGCTG 250
QY 241 GATTGAGGCGCTTCCAGAGTGTATGGTTTTCAGACAGCACCATGATGAATAGAAATGTG 300
DB 251 GATGCGGGGTCTTCAGAGTGTGTGGTTTTCACAGACCATGATGAATAGAAATGTG 310
QY 301 GCCAACAATTTGGTGACAGAGTTCATCGAAGAGTTCGAAGTTTCAAGAGCAGCTCT 360
DB 311 GCCAACAATTTGGTGACAGAGTTCATCGAAGAGTTCGAAGAGTTCGAAGAGCAGCTCT 370
QY 361 ACCTCAGTACATGCCATCAGAAATTTTAAATATATATATATATATATATATATATATAT 420
DB 371 ACCTCAGTACATGCCATCAGAAATTTTAAATATATATATATATATATATATATATATAT 430
QY 421 AATATTCAAGTACTCTCTCATGTTTACATCTACTGATCTTCAAAAGTTGCAGAAATG 480
DB 431 AATATTCAAGTACTCTCTCATGTTTACATCTACTGATCTTCAAAAGTTGCAGAAATG 490
QY 481 ATTGAGAGAGAGATATGATTTCTGTTTCTCTGTTGAGAGCCCATCAGTTTCGATGG 540
DB 491 ATCCGAGAGAGAGATATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 550
QY 541 AGTGAATTCAGAAAGAGTTCGTAAGTGCAGCAACCTCTGAA-TTAAATCCAGCTAAA 600
DB 551 AGTGAATTCAGAAAGAGTTCGTAAGTGCAGCAACCTCTGAA-TTAAATCCAGCTAAA 610
QY 601 CGGCTCTGTCACAAAGACTGGATGAGAAATATATGAAATGCTCATTTATTGCT 660
DB 611 CGGCTCTGTCACAAAGACTGGATGAGAGATTATATGAAAGCGCTCATTTATTGCT 670
QY 661 AAAAGACATTTGATAGATGGTTTACTTGCAGGCT-GGAAATGGCATCTAGCAATG 719
DB 671 AAAAGACATTTGATAGATGGTTTACTTGCAGGCTGGAATATGGGCTATTATGAAATG 730
QY 720 CGAGCTCGAACATAGTGTGGATATAGATGTGGATATTGATGGCTATTGCGAGCAAG 779
DB 731 CGAGCT-GAGCACAGTGTGGATATCGAGTGGCATCATGATTTGGCGCTCGAGAGCAAG 789
QY 780 AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAATATAAATCTTTGGTTG 839
DB 790 AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAATATAAATCTTTGGTTG 849
QY 840 CAATATTGATGGATGTCTCAACATTTGGCAATTGTTGATTCAGGAGACCAAAAGAAAT 899

DB 850 TAAATATTGATGGATGCTCTACCAATGGCCACATTTATGTATCAGAGACCAAAAGAAAT 909
QY 900 AATATCTTATGATGATAAAGATGCTATTGGATAAAGTTTATTAAAGAAAAGTGTATTGA 959
DB 910 AATATCTTATGATGATAAAGATGCTATTGGATAAAGTTTATTAAAGAAAAGCGTATTGA 969
QY 960 GGTGAGGCTTATCTCAGAAAGGGCTGTTTCAAGCAGACGCTGCTCTTTTAAACCTGGA 1019
DB 970 GGTGAGGCTTATCTCAGAAAGGGCTGTTTCAAGCAGACGCTTCCGCCCTTAAAGCTGGA 1029
QY 1020 TTGCAAAATGGAAGTCTGATCAGACAAAGTCTAGCAGTTGTAGATCAATGGAGAAAAGA 1079
DB 1030 CTGTAAAAACAGAAAGTCTGATGTCGATAAAGTCTGCGCACCGTGTAGTGGAGGAGGA 1089
QY 1080 AATGGGCGCTGCTGGGAAAGAGTGCATATCTTTGAAATGAAGTGTCTGATGAAGATG 1139
DB 1090 GATGGGCGCTGCTGGGAAAGAGTGGCTATCTCGGCAATGAAAGTGTCTGATGAAGATG 1149
QY 1140 CTTGAGAGAGTGGGCTTAAGTGGCGCTGCTGATGATGCTGCTTCTACGCCAGAGAGGC 1199
DB 1150 CTTGAGAGAGTGGGCTGAGCGCTGCTTCTGCGGAGCGCTTCTCCGGGGCCCAAGAGC 1209
QY 1200 TGTGTGATACATTTGCAAAATGTAATGCTGCGGTGGTGGTCCCATCCAGAAATTTGAGAGCA 1259
DB 1210 TGTGGGTATACATCTGCAAAATGTCAGCGGTGCGGGGAGGCAATCCGCGAGTTGCGAGCA 1269
QY 1260 CATTTGCTACTAATGCAAAAGTAAATATATATATATATATATATATATATATAT 1305
DB 1270 CATTTGCTACTAATGCAAAAGTAAATATATATATATATATATATATATATATAT 1315

RESULT 13
MMU6215
LOCUS Mus musculus mRNA for CMP-N-acetylneuraminic acid synthetase. 1687 bp mRNA linear ROD 07-AUG-1998
DEFINITION Mus musculus mRNA for CMP-N-acetylneuraminic acid synthetase.
ACCESSION AJ006215.1 GI:3413319
VERSION AJ006215.1
KEYWORDS CMP-N-acetylneuraminic acid synthetase.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 Munster, A.K., Eckhardt, M., Potvin, B., Mühlenhoff, M., Stanley, P. and Gerardy-Schahn, R.
Mammalian cytidine 5'-monophosphate N-acetylneuraminic acid synthetase: a nuclear protein with evolutionarily conserved structural motifs
Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9140-9145 (1998)
98356124
PUBMED 9689047
REFERENCE 2 (bases 1 to 1687)
Gerardy-Schahn, R.
Direct Submission
TITLE Submitted (13-MAY-1998) Gerardy-Schahn R., Institut fuer Medizinische Mikrobiologie, Medizinische Hochschule Hannover, Carl-Neuberg-Str. 1, Hannover, 30625, GERMANY
JOURNAL Location/Qualifiers
FEATURES
source
1..1687
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/cell_line="At-20"
1..1299
/EC_number="2.7.7.43"
/codon_start=1
/product="CMP-N-acetylneuraminic acid synthetase"
/protein_id="CAA06915.1"
/db_xref="GI:3413320"
/db_xref="GOA:O88719"
/db_xref="SPTREMBL:O88719"
/translation="MDALEKAVTSPGPRGPRPPKQSRGAGRGLEKPPHLA
ALVLARGSGKIPKTKRLAGVPLIGWLEALDAGVFQVWVSTDHDEIENVAQF

GAQVHRRSBTSXDSSTSLDAIVEFLNHNHVEVDIVGNIQNTSPCLHPTDLQKVAEMIR
 EGYDSVSVVRHQFWSBISQKQREV-EPLNENPAKPRRQDQWDELYENGSPFPA
 KRLHBMGLYQGGKMAIYEMRAESHDVIDIDWPIAEQVRLFRFGYFGEKKEIKLL
 VCNHIDGLNTHLYVSDOKEXIISYDVKDAIGISLLKSGIEVRLLSERACSKQITLSA
 LKLDCKTEVSVKPLATVDWRKEMGLCKNEVAVLGNEVDEECLKRVGLSAVPADAC
 SGAQKAVGYICKSGRGARFAPAEHIFLLILEKVNNSQK"
 1647..1652

polyA_signal:

ORIGIN

Query Match	77.2%	Score 1007.2;	DB 10;	Length 1687;
Best Local Similarity	87.1%	Prod. No. 1.6e-230;		
Matches 137;	Conservative 4;	Mismatches 157;	Indels 8;	Gaps 3;
QY	1	ATGACCTCGGTGGAGAGGGGGCGCCACCTCGCTCCAAACCCCGGGGGCGACCGCTCC	60	
DB	1	ATGACCGGCTGGAGAGGGGGCGCTCACGTCGGGGGCCCGCCCGCGTGGACGGCGCTCC	60	
QY	61	CGGGGCGGGCGCGAGAGCTGCAGCGCAACTCTCGGGGCGCCAGGGCGGAGGTGTGGAG	120	
DB	61	CGGGGCGGGCCCCCGAGCTCAGCGCA-----GCCGGGGCGGGGGCGCGGCTAGAG	114	
QY	121	AAGCCCGCGCACCTGGCAGCCCTAATCTGSCCGGGGAGGAGCAGCAAGGCAATCCCGCTG	180	
DB	115	AAGCGCGCGCACCTGCGAGCGCTGGTCTGGCGCGCGCGCGCAGCAAGGCAATCCCGCTG	174	
QY	181	AAGAACATTAAGCACCTGGCGGGGGTCCCGCTCATTTGCTGCTGCTCGCGCGCGCGCTG	240	
DB	175	AAGAACATCAAGCGCTGGCGGGGGTTCGCTCATTTGCTGCTGCTCGCGCGCGCGCTG	234	
QY	241	GATTCAGGGGCTTCAGAGGTATGGGTTCGACAGACCATGATGAATTTGAGAAATGTG	300	
DB	235	GATCGGGGGCTTCAGAGGTATGGGTTCGACAGACCATGATGAATTTGAGAAATGTG	294	
QY	301	GCCAAACATTTGGTGCACAGTTCATCGAAGAGTTCCTGAAGTTTCAAAAGACAGCTCT	360	
DB	295	GCCAAACATTTGGTGCACAGTTCATCGAAGAGTTCCTGAAGTTTCAAAAGACAGCTCT	354	
QY	361	ACCTCACTAGATGCCATCATGAATTTCTTAATTAATATGAGSKTGACATTTGAGA	420	
DB	355	ACCTCACTAGAGCCATTTGAGAAATTCCTGAATTAATCAATGAGGTTGACATTTGGGG	414	
QY	421	AATATTCAGCTACTTCATGTTTACATCCCTGCTGATCTTCAAAAAGTTCGAGAAATG	480	
DB	415	AATATTCAGCCACATCTCCAGTTTACATCCCGCTGACCTCCAGAAAGTTCGAGAAATG	474	
QY	481	ATTCGAGAGAGAGGTATGATCTGCTTCTGTTGTGAGACGGCCATCAGTTTCGATGG	540	
DB	475	ATTCGAGAGAGAGGTATGATCTGCTTCTCGTTGTGAGGCGCCATCAGTTTCGATGG	534	
QY	541	AGTGAATTCAGAAAGGAGTTCGTGAAGTACCGAACCTCTGAATTTAAATCCAGCTAA	600	
DB	535	AGTGAATTCAGAAAGGAGTTCGTGAAGTACCGAACCTCTGAAGTTCGAACTTCAGCGAA	594	
QY	601	CGGCTCGTCGACAGCTGGGATGAGAAATTAATGAATAAGTTCATTTTGTCT	660	
DB	595	CGGCTCGTCGACAGCTGGGATGAGAGTTATATGAATAAGGCTCATTTTATTTGTCT	654	
QY	661	AAAAGACATTTGATAGAGATGGTTACTTTCAGGGT-GGAAATGCGCACTACGAAATG	719	
DB	655	AAAAGACATTTGATAGAGATGGTTACTTTCAGGGTGGGAAATGCGCACTATTAAGAAATG	714	
QY	720	CGAGCTGGAAACATAGTGTGATATAGATGATGATATGATGATGATGATGATGATGATG	779	
DB	715	CGAGCT-GGACACAGTGTGATATCGACGTGGACATCGATTTGCCCGATCGCAGAGCAAG	773	
QY	780	AGTATTAAGATATGGCTATTTTCGCAAGAGAGAGCTTAAGAAATAAAACTTTTGGTTTG	839	
DB	774	AGTCTGAGATTTGGCTATTTTGGAAAGAGAGAGCTGAAGGAGATAAAGCTTTTGGTTTG	833	
QY	840	CAATATGATGATGTCTCAACCAATGCGCACTTATGATATGATATGATATGATATGATATG	899	
DB	834	TAATATGATGATGTCTCAACCAATGCGCACTTATGATATGATATGATATGATATGATATG	893	

QY	900	AAATATCTATGATGTAAGATGCTATTGGATAAGTTTATTAAAGAAAAGTGTGATGA	959
DB	894	AAATATCTATGATGTAAGATGCTATTGGATAAGTTTATTAAAGAAAAGTGTGATGA	953
QY	960	GGTGAGGCTAATCTCAGAAAAGGGGCTTTCAAAGCAGACGCTGCTTCTTTAAACCTGGA	1019
DB	954	GGTGAGGCTCATCTCAGAACGGGCTCTCTCCAAGCAGACGCTCTCTCCCTTAAGCTGGA	1013
QY	1020	TTGCAAAATCGAAGTCACTGATCAGACAAAGCTAGAGTGTGTAGATGAATGAGAGAAAGA	1079
DB	1014	CTGTAAACAGAAAGTCACTGATGTCGCGATAGCTGCGCACCGGTGATGAGTGGAGAGGA	1073
QY	1080	AATGGGCTCTGCTGGAAAAGAGTGGCATATCTTTGAAAATGAACTGTGTGATGAAGAGTG	1139
DB	1074	GATGGGCTCTGCTGGAAAAGAGTGGCATATCTCGCAATGAACTGTGTGATGAAGAGTG	1133
QY	1140	CTTGAAGAGAGTGGGCTTAAGTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1199
DB	1134	CTTGAAGAGAGTGGGCTTAAGTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1193
QY	1200	TGTTGGATACATTTGCAAAATGTAATGTTGGTGGCGCTGCTGCTGCTGCTGCTGCTGCTG	1259
DB	1194	TGTGGGTACATCTGCAAAATGTAATGTTGGTGGCGCTGCTGCTGCTGCTGCTGCTGCTG	1253
QY	1260	CATTGCTCTACTAATGAAAAGTAAATAA-TTCATGCCAAAATAAG 1305	
DB	1254	CATTGCTCTACTAATGAAAAGTAAATAA-TTCATGCCAAAATAAG 1299	

RESULT 14

BC016609	1588 bp	mRNA	linear	PRI 03-OCT-2003
Homo sapiens				
LOCUS				
DEFINITION	Homo sapiens cytidine monophosphate N-acetylneuraminic acid synthetase, mRNA (cDNA clone MGC:5424 IMAGE:3448277), complete cds.			
ACCESSION	BC016609			
VERSION	BC016609.1	GI:16741604		
KEYWORDS	MGC.			
SOURCE	Homo sapiens	(human)		
ORGANISM	Homo sapiens			
REFERENCE	1 (bases 1 to 1588)			
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, A., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Brownstein, M.J., Usdin, T.B., Casavant, T.L., Schetz, T.B., Brownstein, M.J., Usdin, T.B., Casavant, T.L., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., Carninci, P., Prange, C., Raha, S., Loquellanc, N.A., Peters, G.J., Scheetz, T.B., Brownstein, M.J., Usdin, T.B., Casavant, T.L., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Villalón, D.K., Murny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sahay, J., Hellon, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.			
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99	(26),	16899-16903 (2002)
MEDLINE	22388257			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 1588)			
AUTHORS	Strausberg, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (31-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2599, USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			

COMMENT Contact: MGC help desk
 Email: gcapbs-femail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCX-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegied, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 3 Row: b Column: 2
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 22027483
 Location/Qualifiers
 1..1588
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:5424 IMAGE:3448277"
 /tissue_type="Placenta, choriocarcinoma"
 /clone_lib="NIH MGC_10"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

 1..1588
 /gene="CMAS"
 /db_xref="LocusID:55907"
 /db_xref="WIM:603316"
 80..871
 /codon_start=1
 /product="CMAS protein"
 /protein_id="AAH16609.1"
 /db_xref="GI:16741605"
 /db_xref="LocusID:55907"
 /translation="MDSVEKGAA"SVSNPRGPRGRPPKLORNSRGQGRGVKEKPH
 LAALLAGGSKGIPLNKIKHLGVLPIGLVLRALDGAFOQSVWVSTDHDEIENAKF
 QFGQVHRSSVFSKSDSTLDALEFLNVHNEVDVGNQATSPCLHPTDLQKVAEM
 IREGDSYFVSVAHRHFRNKEIQGVRETEPLNENKPRPDWDGELYENGSTFY
 FAKHLHMGYLQGGKWAYEMRAHSVDIDVDLDFLEQRVLK"
 215..850
 /note="CTP transf 3; Region: Cytidyltransferase. This
 family consists of two main Cytidyltransferase
 activities: 1) 3-deoxy-manno-octulosonate
 cytidyltransferase,, EC:2.7.7.38 catalysing the
 reaction:- CTP + 3-deoxy-D-manno-octulosonate ==>
 diphosphate + CMP-3-deoxy-D-manno-octulosonate, 2)
 acylneuraminate cytidyltransferase EC:2.7.7.43,,
 catalysing the reaction:- CTP + N-acetylneuraminate ==>
 diphosphate + CMP-N-acetylneuraminate"
 /db_xref="CDD:pfam02348"

 misc_feature
 70.4%; Score 918.4; DB 9; Length: 1588;
 Best Local Similarity 86.0%; Pred. No. 3e-209;
 Matches 1123; Conservative 4; Mismatches 5; Indels 174; Gaps 3

 ORIGIN
 Query Match 70.4%; Score 918.4; DB 9; Length: 1588;
 Best Local Similarity 86.0%; Pred. No. 3e-209;
 Matches 1123; Conservative 4; Mismatches 5; Indels 174; Gaps 3

 QY 1 ATGGACTCGTGGAGAGGGGGCCGCCACCTCGTCTCCACCGCGGGCGACCGTCC 60
 DB 80 ATGGACTCGTGGAGAGGGGGCCGCCACCTCGTCTCCACCGCGGGCGACCGTCC 139
 QY 61 CGGGGCGCGCCCGCCGAAGCTGCAGGCGCAACTCTCGCGGCGCGCAGGCGCGAGTGTGGAG 120
 DB 140 CGGGGCGCGCCCGCCGAAGCTGCAGGCGCAACTCTCGCGGCGCGCAGGCGCGAGTGTGGAG 199
 QY 121 AAGCCCGCGCACTGCGAGCCCTAATTCTGGCCCGGGGAGGCGAGCAAGAGGTCATCCCTGTG 180
 DB 200 AAGCCCGCGCACTGCGAGCCCTAATTCTGGCCCGGGGAGGCGAGCAAGAGGTCATCCCTGTG 259

DB 1167 CATTGGCTACTAATGGAAAGGTTAATAATTCATGCCCAAAATAG 1212

RESULT 15

AX541147
LOCUS AX541147 838 bp DNA linear PAT 23-NOV-2002
DEFINITION Sequence 394 from Patent WO20055700.
ACCESSION AX541147
VERSION AX541147.1 GI:25274513
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Escobedo, J., Garcia, P.D., Kassam, A., Lamson, G., Drmanac, R.,
Crkvenjakov, R., Dickson, M., Drmanac, S., Labat, I., Leshkowitz, D.,
Kita, D., Garcia, V., Jones, W.L., Stache-Crain, B. and Scott, E.M.
Human genes and gene expression products isolated from human
prostate

JOURNAL Patent: WO 02055700-A 394 18-JUL-2002;
Chiron Corporation (US); HYSEQ, INC. (US)

FEATURES
source
1..838
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 48.8%; Score 636.6; DB 6; Length 838;
Best Local Similarity 94.7%; Pred. No. 1e-141;
Matches 715; Conservative 4; Mismatches 26; Indels 10; Gaps 6;
QY 150 GGCCGGGGGAGGAGCAAGGATCCCTCCAGAAACATTAAAGCAGCTGGCGGGGTCCC 209
DB 71 GGCCGGGGGAGGAGCAAGGATCCCTCCAGAAACATTAAAGCAGCTGGCGGGGTCCC 130
QY 210 GTCATATGCTGGTCTGGTGGCGGCTGGATTCAGGGGCTTCCAGAGTGTATGGGT 269
DB 131 GTCATATGCTGGTCTGGTGGCGGCTGGATTCAGGGGCTTCCAGAGTGTATGGGT 190
QY 270 TTCGACAGACCATGATGAATTCAGATGTGGCCAAACATTGGTGGCACAGTTCATCG 329
DB 191 TTCGACAGACCATGATGAATTCAGATGTGGCCAAACATTGGTGGCACAGTTCATCG 250
QY 330 AAGAGTCTCTGAAGTTTCAAAGACAGCTCTACCTCACTAGATGCCATCATAGAAATTC 389
DB 251 AAGAGTCTCTGAAGTTTCAAAGACAGCTCTACCTCACTAGATGCCATCATAGAAATTC 310
QY 390 TAATATATATAGAGGTGACATTTAGGAATATTCAGCTACTCTCTCATGTTTACA 449
DB 311 TAATATATATAGAGGTGACATTTAGGAATATTCAGCTACTCTCTCATGTTTACA 370
QY 450 TCCTACTGATCTTCAAAAGTTCCAGAAATGATTCCAGAAAGAGATATGATCTGKTT 509
DB 371 TCCTACTGATCTTCAAAAGTTCCAGAAATGATTCCAGAAAGAGATATGATCTGKTT 430
QY 510 CTCGTGTGTGAGAGCGGCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAGT 569
DB 431 CTCGTGTGTGAGAGCGGCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAGT 490
QY 570 GACCGACCTCTGAATTTAAATCCAGCTAAACGGCTCTCGACACAGCTGGATGGAGA 629
DB 491 GACCGACCTCTGAATTTAAATCCAGCTAAACGGCTCTCGACACAGCTGGATGGAGA 550
QY 630 ATTATATGAATGGGTCAATTTATTTGCTAAAGACATTTGATAGAGATGGGTACTT 689
DB 551 ATTATATGAATGGGTCAATTTATTTGCTAAAGACATTTGATAGAGATGGGTACTT 610
QY 690 GCAGGGTGG-AAAATGGCTACTACGAAATCGAGCTGGAAACATAGTGTGATATAGATG 748
DB 611 GCAGGGTGGAAATGGCTACTACGAAATCGAGCTGGAAACATAGTGTGATATAGATG 669

QY 749 TGGATATTGATTGGCTATTTCAGAGCAAGAGTATTAGATATGCTATTTTGGCAGG 808
DB 670 TGGATATTGATTGGCTATTTCAGAGCAAGAGTATTAGATATGCTATTTTGG-TRAG 728
QY 809 AGAGCTTAAGGAAATATAAACTTTGGTTTGAATATTGATGATGCTCACCAGTGGCC 868
DB 729 AAGAGCTTAAG--AATAAATTTGGTT--GCAATTTGATGATGCTCACCAGTGGCC 781
QY 869 ACATTATGATATCAGGAGACCAAAAGAAATATA 903
DB 782 CCNTTTTCTGTTTTCAGGAGCAAAAGAAATATA 816

Search completed: May 11, 2004, 17:15:27
Job time : 5315.32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nuc-eic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:29 ; Search time 100.426 Seconds
(without alignments)
7211.402 Million cell updates/sec

Title: US-09-930-440B-3
Perfect score: 1305
Sequence: 1 atggactcgtggaaggg.....ataattcagccaaaatag 1305

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1303.4	99.9	1305	US-09-516-143A-1	Sequence 1, Appli
2	1272.4	97.5	1770	US-09-620-312D-360	Sequence 360, App
3	53.6	4.9	7218	US-08-232-463-14	Sequence 14, Appli
4	50.6	3.9	1830121	US-09-557-884-1	Sequence 1, Appli
5	50.6	3.9	1830121	US-09-543-990A-1	Sequence 1, Appli
6	50.2	3.8	585	US-09-543-681A-1730	Sequence 1730, Ap
7	45.8	3.5	580073	US-08-545-528D-1	Sequence 1, Appli
8	45.6	3.5	684	PCT-US96-05320A-1282	Sequence 1282, Ap
9	42.4	3.2	832	US-09-621-976-2813	Sequence 2813, Ap
10	41.6	3.2	113211	US-09-596-002-40	Sequence 40, Appl
11	41	3.1	1182	US-09-252-991A-6736	Sequence 6736, Ap
12	41	3.1	1848	US-09-252-991A-6953	Sequence 6953, Ap
13	41	3.1	1899	US-09-252-991A-6781	Sequence 6781, Ap
14	40.4	3.1	744	US-09-252-991A-13301	Sequence 13301, A
15	40.4	3.1	783	US-09-252-991A-12981	Sequence 12981, A
16	40.4	3.1	786	US-09-252-991A-12498	Sequence 12498, A
17	40	3.1	640681	US-09-790-988-1	Sequence 1, Appli
18	39.8	3.0	534	US-09-540-236-1677	Sequence 1677, Ap
19	38.6	3.0	3684	US-08-252-991A-65	Sequence 65, Appl
20	38.6	3.0	7527	US-08-252-991A-71	Sequence 71, Appl
21	38.4	2.9	5652	US-09-601-198-75	Sequence 75, Appl
22	38.2	2.9	1134	US-09-252-991A-8176	Sequence 8176, Ap
23	38.2	2.9	1664976	US-09-916-421B-1	Sequence 1, Appli
24	37.8	2.9	1674	US-09-894-844-14	Sequence 14, Appl
25	37.8	2.9	1674	US-09-894-844-15	Sequence 15, Appl
26	37.8	2.9	4411529	US-09-103-840A-1	Sequence 1, Appli
27	37.6	2.9	468	US-09-134-001C-180	Sequence 180, App

37.4	2.9	30001	1	US-08-125-468-1	Sequence 1, Appli
37.4	2.9	30001	2	US-08-474-933-1	Sequence 1, Appli
37.4	2.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
37.2	2.9	474	4	US-09-621-976-18033	Sequence 18033, A
37	2.8	1350	4	US-09-904-615-28	Sequence 28, Appl
36.8	2.8	1807	4	US-09-134-001C-1892	Sequence 1892, Ap
36.8	2.8	1416	4	US-09-328-352-1282	Sequence 1282, Ap
36.6	2.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
36.6	2.8	1131	4	US-09-252-991A-14795	Sequence 14795, A
36.4	2.8	1404	4	US-09-134-001C-398	Sequence 398, App
36.4	2.8	1554	4	US-09-252-991A-14544	Sequence 14544, A
36.4	2.8	1761	4	US-08-252-991A-14820	Sequence 14820, A
36.4	2.8	2911	3	US-08-961-825-5	Sequence 5, Appli
36.4	2.8	2911	3	US-09-480-784-5	Sequence 5, Appli
36.4	2.8	123908	4	US-09-585-858-1	Sequence 1, Appli
36.4	2.8	306	4	US-08-956-171E-4468	Sequence 4468, Ap
36.2	2.8	152331	3	US-09-128-155-16	Sequence 16, Appl
35.8	2.7	942	4	US-09-134-000C-2696	Sequence 2696, Ap

ALIGNMENTS

RESULT 1
US-09-516-143A-1
; Sequence 1, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF505PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; PRIOR FILING DATE: 2003-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1302)
US-09-516-143A-1

Query Match	99.9%	Score 1303.4	DB 4	Length 1305
Best Local Similarity	100.0%	Prod. No. 0		
Matches 1305	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	ATGACTCGGTGGAGAAAGGGGGCGCCACCTTCCTCCAAACCCCGGGGGCGACCGTCC	60	
Db	1	ATGACTCGGTGGAGAAAGGGGGCGCCACCTTCCTCCAAACCCCGGGGGCGACCGTCC	60	
Qy	61	CGGGGGCGGGCGCGAGCTGCAGGGAATCTCTCGCGGGCGCGAGGGGTGCGG	120	
Db	61	CGGGGGCGGGCGCGAGCTGCAGGGAATCTCTCGCGGGCGCGAGGGGTGCGG	120	
Qy	121	AAGCCCGCGACCTTGGCAGCCCTTAATCTGCGCGGGGAGGAGCAGCAAGGCTATCCCTTG	180	
Db	121	AAGCCCGCGACCTTGGCAGCCCTTAATCTGCGCGGGGAGGAGCAGCAAGGCTATCCCTTG	180	
Qy	181	AAGAACATTAAAGACCTTGGCGGGGGTCCCGTCTATGCTGGTCTGGTGGGGCTG	240	
Db	181	AAGAACATTAAAGACCTTGGCGGGGGTCCCGTCTATGCTGGTCTGGTGGGGCTG	240	
Qy	241	GATTTCAGGGGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATG	300	
Db	241	GATTTCAGGGGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATG	300	
Qy	301	GCCAAACATTTGGTGCAAGTTCATCGAAGAGTCTGAAATTTCAAAGACAGCTCT	360	
Db	301	GCCAAACATTTGGTGCAAGTTCATCGAAGAGTCTGAAATTTCAAAGACAGCTCT	360	

D5	585	ATTGGAAGAGGATATGATCTCTGTTTCTCTGTGAGACGCCATCAGTTTCGATGG	644
QY	541	AGTCAAAATTCAGAAAGGAGTTCGTGAAGTGACCGAACTCTCGAATTTAAATCCAGCTAAA	600
D5	645	AGTGAATTCAGAAAGGAGTTCGTGAAGTGACCGAACTCTGAAATTTAAATCCAGCTAAA	704
QY	601	CGGCCTCGTCGACAGACTGGGATGGAGAAATATATGAAAAATGGCTCATTTATTGTCT	660
D5	705	CGGCCTCGTCGACAGACTGGGATGGAGAAATATATGAAAAATGGCTCATTTATTGTCT	764
QY	661	AAAAGACATTTGATAGAGATGGGTACTCTCAGGCTGGAAATATGGCATACTACGAATG	719
D5	765	AAAAGACATTTGATAGAGATGGGTACTCTCAGGCTGGAAATATGGCATACTACGAATG	824
QY	720	CGAGCTGGAACTAGTGTGGATATAGATGTGGATATGATGGCCCTATTGCAGAGCAAG	779
D5	825	CGAGCT-GAACATAGTGTGGATATAGATGTGGATATGATGGCCCTATTGCAGAGCAAG	883
QY	780	AGTATTAAAGATATGGCTATTTTGGCAAGAGAGACTTACGGAAATATAAATTTTGTTTG	839
D5	884	AGTATTAAAGATATGGCTATTTTGGCAAGAGAGACTTACGGAAATATAAATTTTGTTTG	943
QY	840	CAATATTGATGGATGTCTCAACCAATGGCCCACTTATGTATCAGGACACCAAAAAGAAAT	899
D5	944	CAATATTGATGGATGTCTCAACCAATGGCCCACTTATGTATCAGGACACCAAAAAGAAAT	1003
QY	900	AATATCTTATGATGTAAAGATGCTATTGGGAAGATTTATTAAGAAAAAGTGATATGA	959
D5	1004	AATATCTTATGATGTAAAGATGCTATTGGGAAGATTTATTAAGAAAAAGTGATATGA	1063
QY	960	GGTGAGCTTAATCTCAGAAAGGGCCCTGTTCAAGCAGACGCTGTCTTCTTTAAACTGGA	1019
D5	1064	GGTGAGCTTAATCTCAGAAAGGGCCCTGTTCAAGCAGACGCTGTCTTCTTTAAACTGGA	1123
QY	1020	TTSCAAATCGAAGTCAAGTGTATCAGACAAGCTAGCAGTTGTAGTCAATGSGAGAAAGA	1079
D5	1124	TTSCAAATCGAAGTCAAGTGTATCAGACAAGCTAGCAGTTGTAGTCAATGSGAGAAAGA	1183
QY	1080	AATGGGCCCTGTCTGGAAGAAGTGGCATATCTTTGGAATTAAGATGTCTGATCAAGATG	1139
D5	1184	AATGGGCCCTGTCTGGAAGAAGTGGCATATCTTTGGAATTAAGATGTCTGATCAAGATG	1243
QY	1140	CTTGAAGAGAGTGGGCCCTAAGTGGCGCTCTCTGCTGATGCTGTTCTTACGCCAGAGGC	1199
D5	1244	CTTGAAGAGAGTGGGCCCTAAGTGGCGCTCTCTGCTGATGCTGTTCTTACTGCCAGAGGC	1303
QY	1200	TGTTTGGATACATTTGCCAAATGTAATGTTGCGCCGTGGTCCATCCGAGAAATTTCCAGGCA	1259
D5	1304	TGTTTGGATACATTTGCCAAATGTAATGTTGCGCCGTGGTCCATCCGAGAAATTTCCAGGCA	1363
QY	1260	CATTTGCCCTACTAATGGAAAAAGTTAATAATCATGCCAAAAATAG	1305
D5	1364	CATTTGCCCTACTAATGGAAAAAGTTAATAATCATGCCAAAAATAG	1409

RESULT 3

US-08-232-463-14/c

Sequence 14, Application US/08232463

; Patent No. 5670367

1 FACSIM NO. 3373387
1 GENERAL INFORMATION:

1. GENERAL INFORMATION:
 ; APPLICANT: DORNER, F.

APPLICANT: SCHEIFLINGER, F.

APPLICANT: FALKNER, F. G.

APPLICANT: FALNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWL POX VIRUS

;	TITLE OF INVENTION:	REC
:	NUMBER OF SENTENCES:	53

; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:

1. CORRESPONDENCE ADDRESS:
ADDRESSEE: POLICE & PROSECUTOR

ADDRESSEE: Foley & Lardner
ATTENTION: 1000 Wisconsin Road
CITY: 01425

STREET: 1800 Diag

CITY: Alexand

STATE: VA

; COUNTRY: USA

ZIP: 22313-0299

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: ptzgpt-fls
US-08-232-463-14

RESULT 4

US-09-557-884-1

03-09-2004-1
: Sequence 1. Application CS/09557884

Sequence 1, Application
Patent No. 6506581

1. FILE NO. 6506361
2. GENERAL INFORMATION:

GENERAL INFORMATION:
APPLICANT: Fleischmann et al

APPLICANT: Fleischmann et al.

1. TITLE OF INVENTION: The Nucleotide sequence of the *Haemophilus influenzae* pd genome

the *Haemophilus influenzae* type b and type a vaccines

49

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: C8/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 3.9%; Score 50.6; DB 4; Length 1830121;
Best Local Similarity 45.1%; Pred. No. 0.0098;
Matches 228; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
QY 780 AGTATTAAAGATAGGCTATTGTCGAAGAGAGCTTAAGGAAATATAAATTTGGTTG 839
DB 1747282 ATTTCAAGCTTAAGGAATCCGATGCAACAAAATAGAAAATATAAATTTGTCATTAC 1747341
QY 840 CAATATTGATGGATGCTCCACCAATGCCCAATTTATGATCAAGAGACCAAAAGAAAT 899
DB 1747342 TGAATAGACGGGCTACTGATGGAACACTTCATATGATGCCAATGGTGAAGCCAT 1747401
QY 900 AATATCTTATGATGTAAAGATGCTATTGGGATAGTTTATTAAAGAAAAGTGGTATTGA 959
DB 1747402 CAAAAGCTTTCACTACGAGATGGTTTAGGCATATAAATGCTGATGGATCGGATATTCA 1747461
QY 960 GGTGAGGCTAATCTCAGAAAGGCGCTTCAAGACAGACGCTGTCTTCTTTAAACTGGA 1015
DB 1747462 AGTGGCAGTGCTCTTGGTGGCGACTCCCTATTATTTACGTGCTGCGCATGCCGATCTGG 1747521
QY 1020 TTGCAAAATGGAAGTCACTGTATCA---GCAACACTAGCAGTGTAGATGAATGAGAAA 1075
DB 1747522 TATTAATATTCTTCTTGGCAACTTGAAGAAGAACGGCTGTGTTGATCTCATGAA 1747581
QY 1077 AGAAATGGGCGCTGCTGGAAGAAAGAGTGGGATATCTTTGAAATGAACTGTCTGATGAAG 1135
DB 1747582 ACAAGCTGGGCTACCCCGCAGCAACCGCTTATATGGGATGATAGCGTAGATCTCCC 1747641
QY 1137 GTGCTTCAAGAGAGTGGCGCTAGTGGCGCTCTGCTGATGCGCTGTCTTCCACGCCAGAA 1196
DB 1747642 CGCTTTGCAAGCGGTGGAACCTTTCTTTCCTGGCTGATGCGCCCTATTATGTTGAAAAA 1747701
QY 1197 GGCTGTTGGATACATTTGCAAAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGAGA 1256
DB 1747702 TGCTGTGATCATGTACTTTCCACCCATGCGGCAAGAGGGCGCATTCGCTGAATGTCGA 1747761
QY 1257 GCACATTTGCTACTAATGAAAAA 1281

DB 1747762 TATGATTTTACAGCACAGGAAAA 1747786
RESULT 5
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: the Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: C8/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: C8/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 3.9%; Score 50.6; DB 4; Length 1830121;
Best Local Similarity 45.1%; Pred. No. 0.0098;
Matches 228; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
QY 780 AGTATTAAAGATAGGCTATTGTCGAAGAGAGCTTAAGGAAATATAAATTTGGTTG 839
DB 1747282 ATTTCAAGCTTAAGGAATCCGATGCAACAAAATAGAAAATATAAATTTGTCATTAC 1747341
QY 840 CAATATTGATGGATGCTCCACCAATGCCCAATTTATGATCAAGAGACCAAAAGAAAT 899
DB 1747342 TGAATAGACGGGCTACTGATGGAACACTTCATATGATGCCAATGGTGAAGCCAT 1747401
QY 900 AATATCTTATGATGTAAAGATGCTATTGGGATAGTTTATTAAAGAAAAGTGGTATTGA 959
DB 1747402 CAAAAGCTTTCACTACGAGATGGTTTAGGCATATAAATGCTGATGGATCGGATATTCA 1747461
QY 960 GGTGAGGCTAATCTCAGAAAGGCGCTTCAAGACAGACGCTGTCTTCTTTAAACTGGA 1019

Db 1747462 AGTGGCAGTCTTCTGCGGACATCCCTATTTTACGTGCGCATGGCGATCTTGG 1747521
QY 1020 TTGCAAAATGGAAGTCAGTGTATCA---GACAAGCTAGCAGTGTAGATGAAGAGAGAAA 1076
Db 1747522 TATTAAATTTCTTCTTGCGCAACTTGAAAGAGAAACGGCTTGTTGTATCATGAA 1747581
QY 1077 AGAAATGGCCCTGTGCTGGAAGAGAGTGCATATCTTGAAGATGAAGTGTCTGTGAAGA 1136
Db 1747582 ACAAGTGGCTGTCACGCGCAGCAACCGCTTATATTGGCATGATAGCTAGATCCG 1747641
QY 1137 GTGCTTGAAGAGAGTGGGCTAAGTGGGCTCTGCTGTATGCTGTCTAGCCAGAA 1196
Db 1747642 CGCCTTTCGCGGTGGAAGTCTTTTCTGTGGCTGATGCCCTTATTATSTGAAAA 1747701
QY 1197 GCGTGTGTGATACATTTGCAAAATGTAATGTTGGCGCTGTGCGCATCCGAGAAATTCAGA 1256
Db 1747702 TGCTGTGTATGATGATCTTTCCACCCTATGCGGCAAGGCGCATTCGGTGAATGTCGA 1747761
QY 1257 GCACATTTGCCCTACTTAATGGAAAAA 1281
Db 1747762 TATGATTTTACAGGCACAGGAAAA 1747786

RESULT 6

US-09-543-681A-1730
; Sequence 1730, Application US/09543681A
; Patent No. 6605789
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1730
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1730

Query Match 3.8%; Score 50.2; DB 4; Length 585;
Best Local Similarity 47.8%; Pred. No. 5.2e-05;
Matches 207; Conservative 0; Mismatches 223; Indels 3; Gaps 2;
QY 821 AAATAAACTTTTGGTTTGCATATTTGATGATGCTCACCATGSCCAGATTTATGAT 880
Db 92 AAGTCCAATTTGCTGATTTGCGATGTTGATGGTGTGATGCTGACGACCTTATTTATGG 151
QY 881 CAGGAGACCAAAAGAAATAATATCTTATGATGTAAGATGCTATGGGATAAGTTAT 940
Db 152 GCAATAACGGTGAAGATTTGAAGCGTTTAAATGTCGCTGATGCTATGCAATCGTTGT 211
QY 941 TAAAGAAAGTGGTATGAGGTGAGCTAACTCTCAGAAAGGCGCTGTTCAAAGCAGACG 1000
Db 212 TACTTACATCGGTATTCAGGTGCGTATCATACCGGTCGCTCAATCTAAGCTTTTAGAAG 271
QY 2001 TGTCTTTTAAACT--GGATTGCAAAATGGAAGTCACTGTATCA-GACNAGCTAGCAG 1057
Db 272 ATAGAGCCAAAACACTTGGCATTTACATATCTTTCAAGACAGCAATATAAGCTTTTGG 331
QY 2058 TTGTAGATGAATGGAAGAAAGAAATGGGCTGCTGCTGGAAGAGTGGCATATCTTGGAA 1117
Db 332 CGTNTCAACAACTGTTAGATACATAAAACCTTAAACCAAGACAGACAGCTATATGCGG 391
QY 1118 ATGAAGTGTCTGATGAAGAGTGGCTTGAAGAGAGTGGGCTTAAGTGGCGCTCTGCTGAG 1177
Db 392 ATGATCTGATTTGATTTACCGGTAAATGGAAAAAGTAGGACTGTGAGTTGCCGTTGCTGATG 451
QY 1178 CTTGTTCTTCCGCCAGAGAGGCTGTTGGATACATTTGCAATGTAATGTAATGGCGCTGGTG 1237

Db 452 CCATCCATCTACTCACACCTGTCGCCACTATCTTACCCATATTTAGGTGAGCGGTG 511
QY 1238 CCATCCGAGAAAT 1250
Db 512 CAGTACGAGAAAT 524

RESULT 7

US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 3.5%; Score 45.8; DB 4; Length 580073;
Best Local Similarity 48.9%; Pred. No. 0.13;
Matches 113; Conservative 3; Mismatches 115; Indels 0; Gaps 0;
QY 370 GATGCCATCATGAAATTTCTTAATATATATGAGGKTCACATTTAGGAAATATTTCAA 429
Db 564078 GAATACATGATTTATGATATCAATGAAAGCTAAAGTTGGTATGTTATTTCTTCAA 564137
QY 430 GCTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGACAGAAATGATTCGAGAA 489
Db 564138 GATGCTTCTTCATCATGCCCAATTTTGGCAAGTAGTTGCTGGAGATACAGTCAATTA 564197
QY 490 GAAGGATATGATCT 549
Db 564198 AAAGATTATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAACT 564257
QY 550 CAGAAAGGAGTTCGTGAAGTGACCGGACCTCTGAAATTTAAATCCAGCTAAA 600
Db 564258 AGGTTGACGACTACTTGAAGGCAATGTAGAAATTTATGAAAGCAATGCCAAA 564308

RESULT 8

PCT-US96-05320A-1282
; Sequence 1282, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600

; PUBLICATION INFORMATION:

US-09-596-002-40

Query Match 3.1%; Score 41.6; DB 4; Length 119211;

Best Local Similarity 44.6%; Pred. No. 0.79;

Matches 207; Conservative 0; Mismatches 254; Indels 3; Gaps 1;

QY 804 CAAAGAGAGCTTAAGGAAATAAACTTTTGGTTTGCATATATGATGATGCTCACCAG 863
DB 8025 CTACCAAAAGCCAAACATATCAAACTTTTCCCATGATGCTGATTTTTCGGA 80084
QY 864 TGSCACATTTATGATCAGGAGACCAAAAGAAATAATCTATGATGATGATGATG 923
DB 80385 TGGCAATCATTTATAATCTGAAGGCACAAACCAACCAATTTTATGTCAGATGG 80144
QY 924 TATTGGGTAAGTTTATTAAAGAAAGTGGTATTGAGGTGAGGCTAATCTCAGAAAGGC 983
DB 80145 TTTGGGCTTGAAGCGCTAAACAATCAGGCATTATTTGGCCATTATCAGAGTCCGAG 80204
QY 984 CTGTTCAAAGCAGCGCTGCTCTTTTAAACTGG---ATTGCAAAATGGAAGTCACTGT 1040
DB 80205 CTCGCCATGTTGACAGACGCGCCAAAGAGCTTGGCATTTTCACATATTATCAAGGTCA 80264
QY 1041 ATCAGACAGCTAGCAGTTGTAGATGAATGGAAGAAAGAAATGGGCTGTGCTGGAAGA 1100
DB 80265 AGATGATAATGTACTGCTTTGGTGGGCTTCTAAGAGCTTGGCATTTGAGTTATCA 80324
QY 1101 AGTGCCATATCTGAAATGAAGTGTCTGATGAAGAGTCTTTGAAGAGAGTGGGCTTAAG 1160
DB 80325 CTGTGCTTATATCGGTGATGATTTGCTGATCTAAAGCTGTGCTGAGGAGGTTTGG 80384
QY 1161 TGGGCTCTCTGTGATGCTGTCTTCTACCCAGAGGCTGTGATACATTTGCAATG 1220
DB 80385 GATCAGTTACCTAATGGTGTGAGCAGACCCGAGCAGTTTCAGATTATATCAACCAA 80444
QY 1221 TAATGGTGGCTGTGCTGATCCAGAAATTTTCAGAGCACATTT 1264
DB 80445 AACTGGGSGTATGTTGCTGTGAGGAGGTTTGTGAGCTGATTT 80498

RESULT 11

US-09-252-991A-6736

; Sequence 6736, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6736

; LENGTH: 1182

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6736

Query Match

Best Local Similarity 3.1%; Score 41; DB 4; Length 1182;

Matches 100; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

QY 46 CGGGGGCGACCTCCCGGGCGCGCGGAGCTGACGCACTCTCGCGCGGCCAG 105
DB 106 CTGGTGGCGCCCGCGGCTGGCCACCGCGGCTGCGCAAGCTGCTGGTGGCGGTGA 165
QY 106 GCGCGAGGTGTG-GAGAAAGCCCCCGACCTGGCAGCCCTAATTC-TGGCCCGGGAGGCGAG 164
DB 166 GGAACACGAGCGAGAGGCTCCAGGCTTCCGCGCCACGACGACGCTGCGGGTGA 225

QY 165 CAAAGGATTCCTCCCTGAAGAACATTAAGCACCTGGGGGGGTCCTCCCTCATTTGGCTGGGT 224
DB 226 CGACCTGATGTCGCTGGAGCGGACCCACGCTGGCGGCGCTTGAGAGCCCGCCAGGTCCGCGGGT 285
QY 225 C 225
DB 286 C 286

RESULT 12

US-09-252-991A-6953/c

; Sequence 6953, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6953

; LENGTH: 1848

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6953

Query Match

Best Local Similarity 3.1%; Score 41; DB 4; Length 1848;

Matches 100; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

QY 46 CGGGGGCGACCTCCCGGGCGCGGAGCTGACGCACTCTCGCGGCCAG 105
DB 1021 CTGGTGGCGCCCGCGGCTGGCCACCGCGGCTGCGCAAGCTGCTGGCGGTGA 962
QY 106 GCGCGAGGTGTG-GAGAAAGCCCCCGACCTGGCAGCCCTAATTC-TGGCCCGGGAGGCGAG 164
DB 961 GGAACACGAGCGAGAGGCTCGAGGCTTGGCGGCTGCGCAAGCTGCTGGCGGTGA 902

QY 165 CAAAGGATTCCTCCCTGAAGAACATTAAGCACCTGGGGGGGTCCTCCCTCATTTGGCTGGGT 224
DB 901 CGACCTGATGTCGCTGGAGCGGACCAAGCTGCGCGGCTTGAGAGCCCGCCAGGTCCGCGGGT 842
QY 225 C 225
DB 841 C 841

RESULT 13

US-09-252-991A-6781

; Sequence 6781, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6781

; LENGTH: 1899

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6781

Query Match 3.1%; Score 41; DB 4; Length 1899;
Best Local Similarity 55.2%; Pred. No. 0.071;
Matches 100; Conservative 0; Mismatches 80; Indels 1; Gaps 1;
QY 46 CGGGGGGACCGTCCCGGGGCGCGCCGGAAGCTGCAGCGCAACTCTCGGGGGGCGCAG 105
DB 1017 CTGGTGGCCCGCCACGCGGTTCGCCACCGCGGATCGGCAAGCTGCTGGTGGCGGTGGA 1076
QY 106 GGCGAGGTGG-AGAGAGCCCGCGACCTGCGAGCCCTAAATTCGGCCCGGGGAGGCGAG 164
DB 1077 GGAAGAGCGCGGAGAGGCTCCAGGGCCTCGCGGCGCCAAACGACGCGGCTCGCGGGTCGA 1136
QY 165 CAAGGGCATCCCTGTGAGAACATTAAGCACTGCGGGGTCCGGCTCATTTGGCTGGGT 224
DB 1137 CGACTGATCCGCTGAGCGCACACGCTCGCGGCTGGAGGCCCGAGGTCGCGGGGT 1196
QY 225 C 225
DB 1197 C 1197

RESULT 14

US-09-252-991A-13301/c
; Sequence 13301, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13301
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13301

Query Match 3.1%; Score 40.4; DB 4; Length 744;
Best Local Similarity 49.5%; Pred. No. 0.056;
Matches 104; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 33 CGTCTCAACCCGCGGGGCGACCGTCCCGGGGCGCGCCGGAAGCTGCAGCGCAACTC 92
DB 518 CTTGCCCCACCAAGTTCGGGGTTCGAGCGCGGAGGTGCGTCTCGAACAGCATCACTTCG 459
QY 93 TCGGGGGCGGCGAGCGAGGTGTGGAGAACCCCGCCGCACTGGCAGCCCTTAATTCGGC 152
DB 458 GCTCCACCGCGAGCGCGGGGCGATGCGCACGCGCTGTGCTCGCGCGGCTCATGTGCG 399
QY 153 CCGGGAGGCGAGCAAGCAATCCCTGAAGCAATTAAGCACTGGCGGGGTCCCGCT 212
DB 398 CCGGTAGGCGTCTTTCGGGTGCGCCACGCGCACTTGGCCAGGTAGTGTCTGGCCTTCT 339
QY 213 CATTTGGTGGGTCTCTGCGTGGCGGCTTGA 242
DB 338 CGATGGCTTCTTCTTCACACGCGGAGCA 309

RESULT 15

US-09-252-991A-12981/c
; Sequence 12981, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12981
; LENGTH: 783
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12981

Query Match 3.1%; Score 40.4; DB 4; Length 783;
Best Local Similarity 49.5%; Pred. No. 0.058;
Matches 104; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 33 CGTCTCAACCCGCGGGGCGACCGTCCCGGGGCGCGCCGGAAGCTGCAGCGCAACTC 92
DB 580 CTTGCCCCACCAAGTTCGGGGTTCGAGCGCGGAGGTGCGTCTCGCGCGGCTCATGTGCG 521
QY 93 TCGGGGGCGGCGAGCGCGAGGTGTGGAGAACCCCGCCGCACTGGCAGCCCTTAATTCGGC 152
DB 520 GCTCCACCGCGAGCGCGGGGCGATGCGCACGCGCTGTGCTCGCGCGGCTCATGTGCG 461
QY 153 CCGGGAGGCGAGCAAGCAATCCCTGAAGCAATTAAGCACTGGCGGGGTCCCGCT 212
DB 460 CCGGTAGGCGTCTTTCGGGTGCGCCACGCGCACTTGGCCAGGTAGTGTCTGGCCTTCT 401
QY 213 CATTTGGTGGGTCTCTGCGTGGCGGCTTGA 242
DB 400 CGATGGCTTCTTCTTCACACGCGGAGCA 371

Search completed: May 11, 2004, 07:22:40
Job time : 129.426 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 562.652 Seconds
(without alignments)
9853.151 Million cell updates/sec

Title: US-09-930-440B-3

Perfect score: 1305

Sequence: 1 atggactcgtggagaagg.....ataattcatgcacaaataag 1305

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001as.*

5: Geneseq2001bs.*

6: Geneseq2002s.*

7: Geneseq2003as.*

8: Geneseq2003bs.*

9: Geneseq2003cs.*

10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1303.4	99.9	1305	3	AAA50568 Human CMP
2	1303.4	99.9	1305	3	AAA37762 Human gly
3	1303.4	99.9	1305	4	AAB28457 Nucleotid
4	1303.4	99.9	1305	7	AA153992 DNA encod
5	1272.4	97.5	1770	4	AA158475 Human pol
6	1272.4	97.5	1770	8	ADB48450 Novel Hum
7	1272.4	97.5	1800	3	AAH7722 Human can
8	1270.8	97.4	1694	4	AAH17722 Human cDN
9	1270.8	97.4	1751	3	AAH77374 Human ORF
10	1270.8	97.4	1772	3	AAH65392 Human car
11	1229.4	94.2	1803	4	AA160261 Human pol
12	586.8	52.6	738	8	ADB83231 Human cDN
13	547.6	49.6	789	8	ADB83156 Human cDN
14	636.6	48.8	838	6	ABQ89138 Human pro
15	627.4	48.1	745	8	ADB82082 Human cDN
16	625	47.9	819	6	ABQ88853 Human pro
17	620.2	47.5	748	6	ABQ89314 Human pro
18	620.2	47.5	748	8	ADB82257 Human cDN
19	617.4	47.3	708	8	ADB81797 Human cDN
20	562	43.1	686	6	ABQ89876 Human pro
21	562	43.1	686	8	ADB82915 Human cDN
22	555.6	42.6	639	6	ABQ98790 Human ORF
23	538.2	41.2	622	4	AAH06072 Human cDN

24	400.6	30.7	408	5	ACH49109 Human let
25	217	16.6	263	7	ABX44026 Bovine ES
26	181.6	13.9	589	4	Aah11143 Human cDN
27	145.8	11.2	222	6	ABN76702 Human ORF
28	145.4	11.1	205	7	ABX47106 Bovine ES
29	138.2	10.6	478	4	AA112541 Probe #24
30	138.2	10.6	478	4	ABA54246 Human fo
31	138.2	10.6	478	4	ABA43784 Human bre
32	138.2	10.6	478	4	ABA23995 Probe #24
33	138.2	10.6	478	4	AAK27959 Human bon
34	138.2	10.6	478	4	AAK02522 Human bra
35	138.2	10.6	478	4	ABS27544 Human liv
36	138.2	10.6	478	5	AA102446 Probe #24
37	134	10.3	134	4	AA121732 Probe #11
38	134	10.3	134	4	ABA66810 Human fo
39	134	10.3	134	4	ABA48889 Human bre
40	134	10.3	134	4	ABA33872 Probe #12
41	134	10.3	134	4	AAK0362 Human bon
42	134	10.3	134	4	AAK15239 Human bra
43	134	10.3	134	4	ABS40545 Human liv
44	134	10.3	134	5	AA107417 Probe #74
45	96.4	7.4	190	6	ABK35983 cDNA sequ

ALIGNMENTS

RESULT 1

AAA50568

ID AAA50568 standard; cDNA; 1305 BP.

XX AC AAA50568;

XX DT 19-DEC-2000 (first entry)

XX DE Human CMP-sialic acid synthetase cDNA.

XX KW CMP-sialic acid synthetase; human; sialylation; glycoprotein;

XX plasmidogen; transferrin; thyrotropin; Na+, K+-ATPase; ss.

XX CS Homo sapiens.

XX PN WO200052135-A2.

XX PD 08-SEP-2000.

XX PF 01-MAR-2000; 2000WO-US005313.

XX PR 02-MAR-1999; 99US-0122582P.

XX PR 08-DEC-1999; 99US-0169624P.

XX (HUNYA-) HUMAN GENOME SCI INC.

XX (UYJC) UNIV JOHNS HOPKINS.

XX (UYWY-) UNIV WYOMING.

XX PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;

XX WPI; 2000-572178/53.

XX P-PSDB; AAY96100.

XX PT Recombinant production of sialylated glycoproteins using cells in which

XX PT the expression of enzymes, e.g. sialic acid synthetase, involved in the

XX PT sialylation reaction has been altered.

XX PS Disclosure; Page 100-102; 144pp; English.

XX CC The present sequence is that of human cDNA encoding CMP-sialic acid

XX CC synthetase (see AAY96100). This novel gene was identified on the basis of

XX CC homology with the Escherichia coli DNA sequence. The invention provides

XX CC methods and recombinantly engineered cells for producing glycoproteins

XX CC having sialylated oligosaccharides. The methods involve altering the

XX CC expression of enzymes involved in carbohydrate processing. A claimed cell

XX CC producing sialylated glycoprotein at above endogenous levels expresses at

least 1 (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme catalyzing the conversion of UDP-GlcNAc to ManNAc, sialic acid synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid transporter at above endogenous levels. Endogenous N-acetylglucosaminidase activity may be suppressed. A claimed method for manipulating glycoprotein in an insect cell comprises enhancing the expression of 1 of the above enzymes, and a claimed method for producing sialylated glycoproteins involves expressing a heterologous protein (especially plasminogen, transferrin, Na⁺/K⁺-ATPase or thyrotropin) in the insect cell. Yeast, insect, fungal, plant and bacterial host cells can be engineered to produce new forms of sialylated glycoproteins, higher concentrations of sialylated glycoproteins and/or elevated concentrations of donor substrates (e.g. nucleotide sugars) required for sialylation

XX
SQ Sequence 1305 BP; 382 A; 243 C; 350 G; 326 T; 0 U; 4 Other;

Query Match 99.9%; Score 1303.4; DB 3; Length 1305;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACTCGGTGGAGAGGGGGCGCCACCTCCGTCTCCAAACCGCGGGGGCGACCGTCC 60
Db |||||

QY 1 ATGGACTCGGTGGAGAGGGGGCGCCACCTCCGTCTCCAAACCGCGGGGGCGACCGTCC 60
Db |||||

QY 61 CGGGGGCGGGCGGGAGAGTGCAGCGCAACCTCTCGCGCGCGCCAGGCGCGAGGTGTGGAG 120
Db |||||

QY 61 CGGGGGCGGGCGGGAGAGTGCAGCGCAACCTCTCGCGCGCGCCAGGCGCGAGGTGTGGAG 120
Db |||||

QY 121 AAGCCCCCGACCTGGCAGCCCTAATTCCTGCGCGGGAGAGCGAGCAAGGCGATCCCCCTG 180
Db |||||

QY 121 AAGCCCCCGACCTGGCAGCCCTAATTCCTGCGCGGGAGAGCGAGCAAGGCGATCCCCCTG 180
Db |||||

QY 181 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCATCTGGCTGGGTCTCGCTGGCGGGCTG 240
Db |||||

QY 181 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCATCTGGCTGGGTCTCGCTGGCGGGCTG 240
Db |||||

QY 241 GATTCAGGGGCTTCCAGAGTGTATGGTTTCCAGACAGACCATGATGAAATGGAATGTG 300
Db |||||

QY 241 GATTCAGGGGCTTCCAGAGTGTATGGTTTCCAGACAGACCATGATGAAATGGAATGTG 300
Db |||||

QY 301 GCCAACAATTTGGTGCAGAGTTTCATCGAAGAGTTCTGAAAGTTTCAAAGAAGAGTCT 360
Db |||||

QY 301 GCCAACAATTTGGTGCAGAGTTTCATCGAAGAGTTCTGAAAGTTTCAAAGAAGAGTCT 360
Db |||||

QY 361 ACCTCAGTATGTCATCATAGAAATTTCTTAATATATATATATATATATATATATATAT 420
Db |||||

QY 361 ACCTCAGTATGTCATCATAGAAATTTCTTAATATATATATATATATATATATATATAT 420
Db |||||

QY 421 AATATTCAAGTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCAGAAATG 480
Db |||||

QY 421 AATATTCAAGTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCAGAAATG 480
Db |||||

QY 481 ATTTCGAGAGAGAGATATGTTCTGTTCTGTTGTTGAGAGCGGATGATGTTGTTGTTG 540
Db |||||

QY 481 ATTTCGAGAGAGAGATATGTTCTGTTCTGTTGTTGAGAGCGGATGATGTTGTTGTTG 540
Db |||||

QY 541 AGTGAATTCAGAAAGAGTTCGTGAAGTGCACCGACCTCTGAATTAATATATATATATAT 600
Db |||||

QY 541 AGTGAATTCAGAAAGAGTTCGTGAAGTGCACCGACCTCTGAATTAATATATATATATAT 600
Db |||||

QY 601 CGGCTCTGTCGACAGAGTGGGATGAGAAATATATGAAATGGCTCATTTTATTTTGTCT 660
Db |||||

QY 601 CGGCTCTGTCGACAGAGTGGGATGAGAAATATATGAAATGGCTCATTTTATTTTGTCT 660
Db |||||

QY 661 AAGAGACATTTAGATAGATGTTTACTTTCAGGGTGGAAATGGGATCTACGAAATGC 720
Db |||||

QY 661 AAGAGACATTTAGATAGATGTTTACTTTCAGGGTGGAAATGGGATCTACGAAATGC 720
Db |||||

QY 721 GAGCTGGACATAGTGTGATATAGATGATGATATGATGATGATGATGATGATGATGATGAT 780
Db |||||

QY 721 GAGCTGGACATAGTGTGATATAGATGATGATATGATGATGATGATGATGATGATGATGAT 780
Db |||||

QY 781 GTATTAGATATGCTATTTTTCGCAAGAGAGCTTAAGGAATAAATAAACTTTTGGTTTGC 840
Db |||||

QY 781 GTATTAGATATGCTATTTTTCGCAAGAGAGCTTAAGGAATAAATAAACTTTTGGTTTGC 840
Db |||||

QY 841 AATATTGATGATGCTCTCACCATTGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db |||||

QY 841 AATATTGATGATGCTCTCACCATTGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db |||||

QY 901 ATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db |||||

QY 901 ATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db |||||

QY 961 GTGAGGCTAATCTCAGAAAGGGGCTGTTCAAGCAGACGCTGCTTCTTTTAAACTGGAT 1020
Db |||||

QY 961 GTGAGGCTAATCTCAGAAAGGGGCTGTTCAAGCAGACGCTGCTTCTTTTAAACTGGAT 1020
Db |||||

QY 1021 TGCATAATGGAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Db |||||

QY 1021 TGCATAATGGAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Db |||||

QY 1081 ATGGGCTCTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGC 1140
Db |||||

QY 1081 ATGGGCTCTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGC 1140
Db |||||

QY 1141 TTGAAGAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTCTCTCTAGCCCGAGAGGCT 1200
Db |||||

QY 1141 TTGAAGAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTCTCTCTAGCCCGAGAGGCT 1200
Db |||||

QY 1201 GTTGGATACATTTGCAAAATGTAATGTTGGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db |||||

QY 1201 GTTGGATACATTTGCAAAATGTAATGTTGGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db |||||

QY 1261 ATTGGCTCTATGGAAGAGTGAATTAATTCATGCCCAAAATAG 1305
Db |||||

QY 1261 ATTGGCTCTATGGAAGAGTGAATTAATTCATGCCCAAAATAG 1305
Db |||||

RESULT 2
ID AAA37762 standard; DNA; 1305 BP.
XX AAA37762;
XX AC
XX AC
DT 04-DEC-2000 (first entry)
XX
DE Human glycosylation enzyme clone HWL234 coding sequence.
XX
KW Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
KW haematopoietic cell deficiency; blood coagulation disorder; aschma;
KW afibrinogenemia; blood platelet disorder; thrombocytopenia; neoplasia;
KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
KW allergic encephalomyelitis; allergic reaction; organ rejection;
KW graft-versus-host disease; inflammation; hyperproliferative disorder;
KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase; ds.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
CDS 1. .1305
FT /*tag= a
FT /product= "CMP Sialic acid synthetase"
FT /transl_except= (pos:397..399, aa:Xaa)
FT /transl_except= (pos:406..408, aa:Xaa)
FT /transl_except= (pos:439..441, aa:Xaa)
FT /transl_except= (pos:505..507, aa:Xaa)
FT /note= "Xaa= unspecified amino acid"
XX
PN WC2000052136-A2.
XX
PD 03-SEP-2000.
XX
PF 01-MAR-2000; 2000WO-US05325.

[illegible]

Db 1081 ATGGGCTGTCTGGGAAGAGTGGCATATCTTGGAAATGAATGCTGAATGAAGATGC 1140
 QY 1141 TTGAAGAGAGTGGGCTTAAGTGGCTTCTTCTGCTGATGCCCTTCTACGCCCAAGAGCT 1200
 Db 1141 TTGAAGAGAGTGGGCTTAAGTGGGCTTCTTCTGCTGATGCCCTTCTACGCCCAAGAGCT 1200
 QY 1201 GTTGGATACATTGGCAAAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGAGAGCAC 1260
 Db 1201 GTTGGATACATTGGCAAAATGTAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGAGAGCAC 1260
 QY 1261 ATTTCCTACTAATGAAAGATTATATTCATGCGCAAAATAG 1305
 Db 1261 ATTTCCTACTAATGAAAGATTATATTCATGCGCAAAATAG 1305

RESULT 4

AAL53992

ID AAL53992 standard; DNA; 1305 BP.

XX AAL53992;

XX 06-MAR-2003 (first entry)

XX DNA encoding a human CMP-SA synthetase protein.

XX Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
 KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)Nac; sialic acid;
 KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
 KW transporter; sialylated glycoprotein; human; gene; ds.
 XX

CS Homo sapiens.

XX Key Location/Qualifiers

FF 1. .1305
 FT /*tag= a
 FT /product= "Human CMP-SA synthetase protein"
 FT /transl_except= (pos:397..399, aa:His or Tyr)
 FT /transl_except= (pos:406..408, aa:Gly or Val)
 FT /transl_except= (pos:439..441, aa:Pro or Ser)
 FT /transl_except= (pos:505..507, aa:Gly or Val)
 XX

PN US2002142386-A1.

XX 03-OCT-2002.

XX 16-AUG-2001; 2001US-00930440.

XX 02-MAR-1999; 99US-0122582P.

XX 08-DEC-1999; 99US-0169624P.

XX 25-AUG-2000; 2000US-0227579P.

XX (BETE/) BETENBAUGH M J.

XX (LAWR/) LAWRENCE S.

XX (LEBY/) LEE Y C.

XX (COLE/) COLEMAN T A.

XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;

XX WPI; 2003-1C2519/09.

XX P-PSDB; AAO26545.

PT Manipulating glycoprotein production in insect cell, involves enhancing
 PT expression of enzymes involved in carbohydrate processing pathway such as
 PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.

PS Disclosure; Fig 29; 88pp; Engl:sh.

XX The invention relates to a novel method for manipulating glycoprotein
 CC production in an insect cell comprising enhancing expression of an
 CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, ore
 CC catalysing conversion of UDP-GlcNAc to mannose (Man)Nac, sialic acid
 CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)

CC synthetase or CMP-SA transporter, where the expression of each enzyme is
 CC enhanced to above endogenous levels. The novel method is useful for
 CC manipulating glycoprotein production in an insect cell. Further methods
 CC of the invention are useful for producing sialylated glycoprotein. The
 CC sialylated glycoprotein produced by the above mentioned methods are
 CC useful as pharmaceutical compositions, vaccines, diagnostics and
 CC therapeutics. This polynucleotide sequence represents the DNA encoding a
 CC human CMP-SA synthetase protein of the invention
 XX
 SQ Sequence 1305 BP; 382 A; 243 C; 350 G; 326 T; 0 U; 4 Other;

Query Match 99.9%; Score 1303.4; DB 7; Length 1305;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACTCGTGGAGAGGGGGCGCCACCTCCGTCTCCAAACCGCGGGGGCGACCGTCC 60
 Db 1 ATGGACTCGTGGAGAGGGGGCGCCACCTCCGTCTCCAAACCGCGGGGGCGACCGTCC 60
 QY 61 CGGGGGCGCGCGCGAAGCTGAGCGCACTCTCGCGCGCGCGCGCGCGCGCGCGCGAG 120
 Db 61 CGGGGGCGCGCGCGAAGCTGAGCGCACTCTCGCGCGCGCGCGCGCGCGCGCGCGAG 120
 QY 121 ARGCCCCCGCACCTGCGCGCCCTAATTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 Db 121 ARGCCCCCGCACCTGCGCGCCCTAATTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 QY 181 AAGAACATTAAAGCACCTGCGCGGGGTCCCGCTCATTTGCTGGGTGCGTGGCGCGCG 240
 Db 181 AAGAACATTAAAGCACCTGCGCGGGGTCCCGCTCATTTGCTGGGTGCGTGGCGCGCG 240
 QY 241 GATTAGGGGCGCTTCAGAGTGATGGGTTTCACAGACCATGATGTAATGAGAATGTG 300
 Db 241 GATTAGGGGCGCTTCAGAGTGATGGGTTTCACAGACCATGATGTAATGAGAATGTG 300
 QY 301 GCCAAACAAATTTGGTGCACAAAGTTTCATCGAAGAGTTCTCAAGTTTCAAAGACAGCTCT 360
 Db 301 GCCAAACAAATTTGGTGCACAAAGTTTCATCGAAGAGTTCTCAAGTTTCAAAGACAGCTCT 360
 QY 361 ACCTCACTAGATGCCATCATAGAATTTCTTAATATATATATATATATATATATATAT 420
 Db 361 ACCTCACTAGATGCCATCATAGAATTTCTTAATATATATATATATATATATATATAT 420
 QY 421 AATATTCAGCTACTTCTTCATGATCTTACATCTTACATCTTACATCTTACATCTTACAT 480
 Db 421 AATATTCAGCTACTTCTTCATGATCTTACATCTTACATCTTACATCTTACATCTTACAT 480
 QY 481 ATTGGAAGAAGAGGATATGATGATCTGKTTTCTGTTGTGAGAGCGCATCTAGTTTCGATGG 540
 Db 481 ATTGGAAGAAGAGGATATGATGATCTGKTTTCTGTTGTGAGAGCGCATCTAGTTTCGATGG 540
 QY 541 AGTGAATTCAGAAAGAGGTTCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 600
 Db 541 AGTGAATTCAGAAAGAGGTTCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 600
 QY 601 CGGCTCTCTGCAACAGACTGGGATGAGAAATATATATGAAATGGCTCATTTATTTTGTCT 660
 Db 601 CGGCTCTCTGCAACAGACTGGGATGAGAAATATATATGAAATGGCTCATTTATTTTGTCT 660
 QY 661 AAAAGACATTTGATAGAGATGGGTTACTTTCAGGGGTGGAATATGAGCATCTAGCAATGC 720
 Db 661 AAAAGACATTTGATAGAGATGGGTTACTTTCAGGGGTGGAATATGAGCATCTAGCAATGC 720
 QY 721 GAGCTGGAAACATAGTGTGGATATAGATGGAATATGATGGAATATGATGGAATATGATGGA 780
 Db 721 GAGCTGGAAACATAGTGTGGATATAGATGGAATATGATGGAATATGATGGAATATGATGGA 780
 QY 781 GTATTAGATATGGCTATTTTTCGAAAGAGAGCTTAAGGAATATAAATTTTGGTTTTC 840
 Db 781 GTATTAGATATGGCTATTTTTCGAAAGAGAGCTTAAGGAATATAAATTTTGGTTTTC 840
 QY 841 AATATTGATGGATGTCCTACCAATGGCCACATTTATGATATCAGGAGACCAAAAGAAATA 900

841 AATATTGATGATGCTCACCATAATGGCCACATTTATGATCAGGAGACCAAAAAAGAAATA 900
 901 ATATCTTATGATGATTAAGAGATGCTATGGGATAGTTTATTAAGAAAGAGTATTGAG 960
 902 ATATCTTATGATGATTAAGAGATGCTATGGGATAGTTTATTAAGAAAGAGTATTGAG 960
 961 GTGAGGCTAATCTCAGAAAAGGCGCTGTTCAAAGCAGACGCTGTCTTTTAAACTGGAT 1020
 962 GTGAGGCTAATCTCAGAAAAGGCGCTGTTCAAAGCAGACGCTGTCTTTTAAACTGGAT 1020
 1021 TGCAAAATGGAGTCAAGTATCAGACAGCTAGCAGTCTAGATGATGAGAAAGAA 1080
 1022 TGCAAAATGGAGTCAAGTATCAGACAGCTAGCAGTCTAGATGATGAGAAAGAA 1080
 1081 ATGGGCGCTGTCTCGAAAGAGTGGCAATCTTGGAAATCAAGTCTGTGATGAAGAGTGC 1140
 1082 ATGGGCGCTGTCTCGAAAGAGTGGCAATCTTGGAAATCAAGTCTGTGATGAAGAGTGC 1140
 1141 TTGAAGAGATGGCGCTTAAGTGGCGCTCTCTGCTGATGCTCTTCTAGCCAGAGGCT 1200
 1142 TTGAAGAGATGGCGCTTAAGTGGCGCTCTCTGCTGATGCTCTTCTAGCCAGAGGCT 1200
 1201 GTTGATACATTTGCCAATGTAATGGTGGCGCTGTGCTCCGAGATTTGCCAGAGCAC 1260
 1202 GTTGATACATTTGCCAATGTAATGGTGGCGCTGTGCTCCGAGATTTGCCAGAGCAC 1260
 1261 ATTTGCTACTAATGGAAGAAAGTAAATTAATTCATGCCAATAATAG 1305
 1262 ATTTGCTACTAATGGAAGAAAGTAAATTAATTCATGCCAATAATAG 1305

RESULT 5

AA158475
 ID AA158475 standard; cDNA; 1770 BP.
 AC AA158475;
 XX
 22-OCT-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 678.
 XX
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.

DR P-ESDB; AAM39319.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 as central nervous system injuries.
 XX
 PS Claim 1; SEQ ID NO 678; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157738-AA161369) and the
 encoded polypeptides (AAM38642-AAM42213) with neotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 in gene therapy. A composition containing a polypeptide or polynucleotide
 of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC -calicised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 SQ Sequence 1770 BP; 520 A; 311 C; 459 G; 480 T; 0 U; 0 Other;

Query Match 97.5%; Score 1272.4; DB 4; Length 1770;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY 1 ATGCACTCGGTGGAGAGGGGGCGCCACCTCCGCTTCCAAACCCCGGGGGGACCGTCC 60
 DB 105 ATGCACTCGGTGGAGAGGGGGCGCCACCTCCGCTTCCAAACCCCGGGGGGACCGTCC 164
 QY 61 OGGGGCGGGCGCGAAGCTCGACGGCAACTCTCGCGCGGGCGCGGGCGGGTGGAG 120
 DB 165 OGGGGCGGGCGCGAAGCTCGACGGCAACTCTCGCGCGGGCGCGGGCGGGTGGAG 224
 QY 121 AAGCCCCCGCACCTGGCAGCCCTTAATCTGCGCCCGGGGAGGAGCAAGGATCCCGCTG 180
 DB 225 AAGCCCCCGCACCTGGCAGCCCTTAATCTGCGCCCGGGGAGGAGCAAGGATCCCGCTG 284
 QY 181 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCATATGCTGGTCTCTCGTGGCGCCCTG 240
 DB 285 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCATATGCTGGTCTCTCGTGGCGCCCTG 344
 QY 241 GATTTCAGGGGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAATGAGAAATG 300
 DB 345 GATTTCAGGGGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAATGAGAAATG 404
 QY 301 GCCAAACAATTTGGTGACAAAGTTTCATCGAAGAAAGTTTCTGAAGTTTCAAAAAGAGCAGCTCT 360
 DB 405 GCCAAACAATTTGGTGACAAAGTTTCATCGAAGAAAGTTTCTGAAGTTTCAAAAAGAGCAGCTCT 464
 QY 361 ACCTCAGTAGATGCCATCAGAAATTTCTTAATATATATATGAGGTGACATTTGTAGGA 420
 DB 465 ACCTCAGTAGATGCCATCAGAAATTTCTTAATATATATATGAGGTGACATTTGTAGGA 524
 QY 421 AATATTCAGCTACTCTCTCATGTTTACATCTCTACTGATCTTCAAAAAGTTTCAGAAATG 480
 DB 525 AATATTCAGCTACTCTCTCATGTTTACATCTCTACTGATCTTCAAAAAGTTTCAGAAATG 584
 QY 481 ATTTCGAGAAAGAGATGATGATTCGKTTTCTCTTTCTGTGTGAGACCCCATCATGTTGATGG 540
 DB 585 ATTTCGAGAAAGAGATGATGATTCGKTTTCTCTTTCTGTGTGAGACCCCATCATGTTGATGG 644
 QY 541 AGTGAATTCAGAAAGAGTTCGTGAAGTACCCGACCTCTGATTTTAAATCCAGCTAAA 600
 DB 645 AGTGAATTCAGAAAGAGTTCGTGAAGTACCCGACCTCTGATTTTAAATCCAGCTAAA 704
 QY 601 CGGCTTCGTGACAGACTCGGATGGAGAAATTAATGAAATGGCTCATTTTATTTTGGCT 660
 DB 705 CGGCTTCGTGACAGACTCGGATGGAGAAATTAATGAAATGGCTCATTTTATTTTGGCT 764

Db 600 AATCAGAGAGAGGATATGATTTCTGTTCTGTTGAGAGCCATCAGTTTCGATG 659
 QY 541 AGTGAATTCAGAAAGAGTTCGTGAAGTGAACCACTCTGAATTAATTAATCCAGCTAA 600
 Db 660 AGTGAATTCAGAAAGAGTTCGTGAAGTGAACCACTCTGAATTAATTAATCCAGCTAA 719
 QY 601 CGGCTCTGTCGACAGACGTGGATGAGAAATTAATGAATAATGGCTCAATTTATTTGCT 660
 Db 720 CGGCTCTGTCGACAGACGTGGATGAGAAATTAATGAATAATGGCTCAATTTATTTGCT 779
 QY 661 AAAGACATTTGATAGAGTGGTCTACTTGCAGGTGG-AAAATGCACTACTACGAAATG 719
 Db 780 AAAGACATTTGATAGAGTGGTCTACTTGCAGGTGGAAATGCACTACTACGAAATG 839
 QY 720 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATGATGGCTTATTCAGAGCAAG 779
 Db 840 CGAGCT-GAACAATAGTGTGGATATAGATGTGGATATGATGGCTTATTCAGAGCAAG 898
 QY 780 AGTATTAAGATATGCTATTTGGCAAGAGAGCTTAAGGAAATTAATACTTTTGTGTTG 839
 Db 899 AGTATTAAGATATGCTATTTGGCAAGAGAGCTTAAGGAAATTAATACTTTTGTGTTG 958
 QY 840 CAATATGATGATGCTCACCAGTGGCCACATTTATGATATCAGAGAGCAAAAGAAAT 899
 Db 959 CAATATGATGATGCTCACCAGTGGCCACATTTATGATATCAGAGAGCAAAAGAAAT 1018
 QY 900 AATATCTTATGATGATAAGATGCTATTTGGGATAAGTTTATTAAGAAAGTGGTATGGA 959
 Db 1019 AATATCTTATGATGATAAGATGCTATTTGGGATAAGTTTATTAAGAAAGTGGTATGGA 1078
 QY 960 GGTGAGCTTATCTCAGAAAGGCTGTTCAAGCGAGCGTCTCTTTTAAACTGGA 1019
 Db 1079 GGTGAGCTTATCTCAGAAAGGCTGTTCAAGCGAGCGTCTCTTTTAAACTGGA 1138
 QY 1020 TTGCAAAATGGAAGTCAGTGTATCAGCAAGCTAGCAGTTGATGATGAATGAGAAAGA 1079
 Db 1139 TTGCAAAATGGAAGTCAGTGTATCAGCAAGCTAGCAGTTGATGATGAATGAGAAAGA 1198
 QY 1080 AATGGCTGCTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTG 1139
 Db 1199 AATGGCTGCTGCTGGAAGAGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTG 1258
 QY 1140 CTTGAGAGAGTGGCCCTAAGTGGCGCTCTGCTGATGCTGCTGATGCTGCTGATGCTG 1199
 Db 1259 CTTGAGAGAGTGGCCCTAAGTGGCGCTCTGCTGATGCTGCTGATGCTGCTGATGCTG 1318
 QY 1200 TGTGATACATTTGCAAAATGTAATGGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259
 Db 1319 TGTGATACATTTGCAAAATGTAATGGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1378
 QY 1260 CATTGCTCTAATGGAAGTTTAATTAATTCATGCTGCAAAATAG 1305
 Db 1379 CATTGCTCTAATGGAAGTTTAATTAATTCATGCTGCAAAATAG 1424

RESULT 8
 AAH17722
 ID AAH17722 standard; cDNA, 1694 BP.
 XX
 AC AAH17722;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:17324.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.
 PF 29-JUL-1999; 99JP-00248036.
 XX 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WEL; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 PT
 PT
 XX
 PS Claim 8; SEQ ID NO 17324; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH3628 and AAH3633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 CC
 XX
 SQ Sequence 1694 BP; 494 A; 295 C; 433 G; 472 T; 0 U; 0 Other;
 Query Match 97.4%; Score 1270.8; DB 4; Length 1694;
 Best Local Similarity 99.1%; Pred. No. 0;
 Matches 1294; Conservative 4; Mismatches 6; Indels 2; Gaps 2;
 QY 1 ATGGACTCGTGGAGAGGGGGCGCCACCTCCGCTCCACCCCGGGGGGACCGCTCC 60
 Db 41 ATGGACTCGTGGAGAGGGGGCGCCACCTCCGCTCCACCCCGGGGGGACCGCTCC 100
 QY 61 CGGGGCGCGCCCGGAGCTGCAGCGCACTCTCGCGGGCGCAGGGCGAGGTGTGGAG 120
 Db 101 CGGGGCGCGCCCGGAGCTGCAGCGCACTCTCGCGGGCGCAGGGCGAGGTGTGGAG 160
 QY 121 AAGCCCCGACCTCGAGCCCTTAATTCCTGGCCCGGGGAGGAGCAAGGCAATCCCCCTG 180
 Db 161 AAGCCCCGACCTCGAGCCCTTAATTCCTGGCCCGGGGAGGAGCAAGGCAATCCCCCTG 220
 QY 181 AAGAACATTAAGCACTCGCGGGGTCCCGCTCATTTGGCTGGGTCTCGTGGCGCCCTG 240
 Db 221 AAGAACATTAAGCACTCGCGGGGTCCCGCTCATTTGGCTGGGTCTCGTGGCGCCCTG 280
 QY 241 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAGATGTG 300
 Db 281 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCAGAGATGTG 340
 QY 301 GCCAAACAATTTGGTGTGCAAGTTTCATCGAAGAGTTCGAGAGTTTCAAAAGACAGCTCT 360

QY 1 ATGGACTGGTGGAGAGGGGGCCGACCTCGCTCCCAACCCGCGGGGGCGACCGTCC 60
Db 76 ATGGACTGGTGGAGAGGGGGCCGACCTCGCTCCCAACCCGCGGGGGCGACCGTCC 135
QY 61 CGGGCCCGCCCGCAGAGCTGAGCGCACTCTCGGGGGGCGAGCGCGAGGTGTGGAG 120
Db 136 CGGGCCCGCCCGCAGAGCTGAGCGCACTCTCGGGGGGCGAGCGCGAGGTGTGGAG 195
QY 121 AAGCCGCCGCACTGGCGACCCCTAAATCTTGGCCCGGGAGGAGCAGCAAAAGGCATCCCCCTG 180
Db 196 AAGCCGCCGCACTGGCGACCCCTAAATCTTGGCCCGGGAGGAGCAGCAAAAGGCATCCCCCTG 255
QY 181 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCATTTGGCTGGGTCTGGTGGCGGCCCTG 240
Db 256 AAGAACATTAAGCACTGGCGGGGGTCCCGCTCATTTGGCTGGGTCTGGTGGCGGCCCTG 315
QY 241 GATTGAGGGCCCTTCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAGATGG 300
Db 316 GATTGAGGGCCCTTCAGAGTGTATGGGTTCGACAGACCATGATGAATTCAGAGATGG 375
QY 301 GCCAAACAATTTGGTGCACAGTTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT 360
Db 376 GCCAAACAATTTGGTGCACAGTTTCATCGAAGAAGTTCTGAAGTTTCAAAAGACAGCTCT 435
QY 361 ACCTCACTAGATGCCATCATAGAAATTTCTTAATTAATTAATGAGGKTGACATTTAGGA 420
Db 436 ACCTCACTAGATGCCATCATAGAAATTTCTTAATTAATTAATGAGGKTGACATTTAGGA 495
QY 421 AATATTCAAGTACTCTTCATGTTTACATCCCTACTGATCTTCAAAAGTTCCAGAAATG 480
Db 496 AATATTCAAGTACTCTTCATGTTTACATCCCTACTGATCTTCAAAAGTTCCAGAAATG 555
QY 481 ACTCGAAGAAGGATATGATTCGTCTTCTGTTGTGAGCGCCATCAGTTTCGATGG 540
Db 556 ACTCGAAGAAGGATATGATTCGTCTTCTGTTGTGAGCGCCATCAGTTTCGATGG 615
QY 541 AGTGAAATTCAGAAAGGATTCGTGAGTGCAGGACCGAACTCTGAATTTAAATCCAGCTAA 600
Db 616 AGTGAAATTCAGAAAGGATTCGTGAGTGCAGGACCGAACTCTGAATTTAAATCCAGCTAA 675
QY 601 CGGCTCTGTCGACAAAGCTGGATGGAGAAATATATGAAATGGCTCATTTTATTTGCT 660
Db 676 CGGCTCTGTCGACAAAGCTGGATGGAGAAATATATGAAATGGCTCATTTTATTTGCT 735
QY 661 AAAAGA CATTTCATAGAGATGGTTTACTTTCAGAGGTGG-AAAATGGCATCTACGAAATG 719
Db 736 AAAAGACATTTGATAGAGATGGTTTACTTTCAGAGGTGGAAATGCGCATCTACGAAATG 795
QY 720 CGAGCTGAA CATTGTTGGATATAGATGTGATATTCATTGGCCCTATTGAGAGCAAG 779
Db 796 CGAGCT-GAATGATGTGGATATAGATGTGATATTCATTGGCCCTATTGAGAGCAAG 854
QY 780 AGTATTAGATATGGCTATTTTGGCAAAGAGAGCTTAAGGAAATAAACITTTTGGTTG 839
Db 855 AGTATTAGATATGGCTATTTTGGCAAAGAGAGCTTAAGGAAATAAACITTTTGGTTG 914
QY 840 CAATATTGATGATGTCACCAATGGCCCATTTATGATATCAGGAGCCAAAGAAAT 899
Db 915 CAATATTGATGATGTCACCAATGGCCCATTTATGATATCAGGAGCCAAAGAAAT 974
QY 900 AATATCTTATGATGTAAGAGATGCTATTGGGATAGTTTATTAAGAAAGTGGTATTGA 959
Db 975 AATATCTTATGATGTAAGAGATGCTATTGGGATAGTTTATTAAGAAAGTGGTATTGA 1034
QY 960 GGTGAGGCTAATCTCAGAAAGGGCTGTTCAAAGCAGACGGTGTCTTCTTTAAACTGGA 1019
Db 1035 GGTGAGGCTAATCTCAGAAAGGGCTGTTCAAAGCAGACGGTGTCTTCTTTAAACTGGA 1094
QY 120 TGCATAATGGAAGTCACTGTAACAGCAAGCTAGCAGTTGATGATGAATCGAGAAAGA 1079
Db 1095 TTGCAAAATGGAAGTCACTGTAACAGCAAGCTAGCAGTTGATGATGAATCGAGAAAGA 1154
QY 1080 AATGGGCTGTGCTGGAAAGAGTGGCCATATCTTGGAAATGAAGTCTCTGATGAAGATG 1139

Db 1155 AATGGGCTGTGCTGGAAAGAGTGGCATATCTTGGAAATGAGTGTCTGATGAAGATG 1214
QY 1140 CTTGAAGACAGTGGGCTTAAGTGGCGCTCTCTGCTGATGCTGTTCCTAGCCCGAGAGGC 1199
Db 1215 CTTGAAGACAGTGGGCTTAAGTGGCGCTCTCTGCTGATGCTGTTCCTAGCCCGAGAGGC 1274
QY 1200 TGTGATACATTTGCAAAATGTAATGGTGGCGCTGGTGGCCATCCGAGAAATTTGCAGAGCA 1259
Db 1275 TGTGATACATTTGCAAAATGTAATGGTGGCGCTGGTGGCCATCCGAGAAATTTGCAGAGCA 1334
QY 1260 CATTTGCCCTACTTAATGGAAGAGTAAATTAATTCATGCCAAATATAG 1305
Db 1335 CATTTGCCCTACTTAATGGAAGAGTAAATTAATTCATGCCAAATATAG 1380

RESULT 10
AAC65392
ID AAC65392 standard; cDNA; 1772 bp.
XX
AC AAC65392;
XX
DT 13-FEB-2001 (first entry)
XX
DE Human carbohydrate-modifying enzyme cDNA Incyte ID No: 000422CB1.
XX
KW Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW immunosuppressive; anti-HIV; anti-inflammatory; antianemic;
KW antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW nephrotropic; angiot; thyromimetic; neuroprotective; osteopathic;
KW antarthritic; antiproliferative; uropathic; ophthalmological;
KW dermatological; antitumor; cytostatic; virucide; antibacterial;
KW fungicide; procoagulant; tranquilizer; vulnery; diabetes;
KW autoimmune disorder; inflammatory disorder; infection; ss.
XX
OS Homo sapiens.
XX
FN WO200063351-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-US010882.
XX
PR 21-APR-1999; 99US-0130383P.
XX
PA {INCYTE} INCYTE GENOMICS INC.
XX
PI Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;
XX
XX WPI; 2000-672729/65.
DR P-PSDB; AAB28673.
XX
PT Novel carbohydrate modifying enzyme polypeptides and polynucleotides for
PT diagnosis, treatment, and prevention of carbohydrate metabolism
PT disorders, autoimmune/inflammatory disorders, and cancer.
XX
PS Claim 4; Page 72; 75pp; English.
XX
CC The present cDNA sequence encodes a human carbohydrate-modifying enzyme
CC (CME). CME polynucleotides and polypeptides are useful for treating and
CC diagnosing diseases associated with CME such as diabetes,
CC autoimmune/inflammatory disorders such as AIDS, Addison's disease, adult
CC respiratory distress syndrome, allergies, anaemia, asthma,
CC atherosclerosis, autoimmune thyroiditis, bronchitis, cholecystitis,
CC contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC Reiter's syndrome, arthritis, scleroderma, Sjogren's syndrome, systemic
CC lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC complications of cancer, haemodialysis, and extracorporeal circulation,
CC viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC trauma, or cancer. CME, or its catalytic or immunogenic fragment, is

CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AA138642-AA142213) with neurotrophic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 SQ Sequence 1803 BP; 483 A; 461 C; 330 G; 529 T; 0 U; 0 Other;

Query Match 94.2%; Score 1229.4; DS 4; Length 1803;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1274; Conservative 4; Mismatches 5; Indels 4; Gaps 4;

QY 1 ATGSACTCGTGGAGAGGGGCGCCACCTCGCTCCCAACCCGCGGGGCGACCGTCC 60
 Db 1699 ATGSACTCGTGGAGAGGGGCGCCACCTCGCTCCCAACCCGCGGGGCGACCGTCC 1640

QY 61 CGGSCCGGCGCCCGAAGCTGCAGGCGCACTCTCGCGGGCGGCGAGGCGAGGTGGAG 120
 Db 1639 CGGSCCGGCGCCCGAAGCTGCAGGCGCACTCTCGCGGGCGGCGAGGCGAGGTGGAG 1580

QY 121 AAGCCCGCGACCTCGGAGCCCTTATCTCGCGCGGGGCGAGGCGCAAGGCGATCCCGCTG 180
 Db 1579 AAGCCCGCGACCTCGGAGCCCTTATCTCGCGCGGGGCGAGGCGCAAGGCGATCCCGCTG 1520

QY 181 AAGAACATTAAACACTCGCGGGGCTCGCGCTCATTTGGTGGGTCTCGTGGCGGCGCTG 240
 Db 1519 AAGAACATTAAACACTCGCGGGGCTCGCGCTCATTTGGTGGGTCTCGTGGCGGCGCTG 1460

QY 241 GATTTCAGGGGCTCCAGAGTGTATGGGTTTCGACAGACCATGTA-GAAATTGAGATGTG 300
 Db 1459 GATTTCAGGGGCTCCAGAGTGTATGGGTTTCGACAGACCATGTA-GAAATTGAGATGTG 1400

QY 301 GCGAAGCAATTGTGCGACAGTGTGATCGAAGAGTCTGGAAGTTCGAAAGACAGCTCT 360
 Db 1399 GCGAAGCAATTGTGCGACAGTGTGATCGAAGAGTCTGGAAGTTCGAAAGACAGCTCT 1340

QY 361 ACCTCAGTGTGATGATCATGATGATTTCTTAATTATATATGAGGKTGACATCTGAGA 420
 Db 1339 ACCTCAGTGTGATGATCATGATGATTTCTTAATTATATATGAGGKTGACATCTGAGA 1280

QY 421 AATATCAAGCTACTCTCCATGTTTACATCTACTGATCTTCAAAAAGTTCGAGAAATG 480
 Db 1279 AATATCAAGCTACTCTCCATGTTTACATCTACTGATCTTCAAAAAGTTCGAGAAATG 1220

QY 481 ATTTCGAGAGAGGATGATGATCTGKTTCTCTGTTGTTGAGACCCATCAGTTTCGATGG 540
 Db 1219 ATTTCGAGAGAGGATGATGATCTGKTTCTCTGTTGTTGAGACCCATCAGTTTCGATGG 1160

QY 541 AGTCGAAATTGAGAAAGGATGTTGTTGAGTGCAGGACCTCTGAATTTAAATCCAGCTAAA 600
 Db 1159 AGTCGAAATTGAGAAAGGATGTTGTTGAGTGCAGGACCTCTGAATTTAAATCCAGCTAAA 1100

QY 601 CGGCTCTGCGACAGACTGGGATCGAGAAATPATATGAAATGGCTCATTTTATTTGCT 660
 Db 1099 CGGCTCTGCGACAGACTGGGATCGAGAAATPATATGAAATGGCTCATTTTATTTGCT 1040

QY 661 AAGACACATTTGATAGATGGGTACTGTCAGGGTGG-AAATGGGCTACTACGAAATG 719
 Db 1039 AAGACACATTTGATAGATGGGTACTGTCAGGGTGGAAATGGGCTACTACGAAATG 980

QY 720 CGAGCTGGAACATAGTGTGGATACAGATGTGATATTGATTGGGCTATTGCGAGCAAG 779
 Db 379 CGAGCT-GAACATAGTGTGGATACAGATGTGATATTGATTGGGCTATTGCGAGCAAG 921

QY 780 AGTATTAGATATGGCTATTGTTGGCAAGAGAGCTTAAGGAATAAATCTTTGTTGTTG 839
 Db 920 AGTATTAGATATGGCTATTGTTGGCAAGAGAGCTTAAGGAATAAATCTTTGTTGTTG 861

QY 840 CAATATTGATGATGCTCACCAGTGGCCAGCTTATGTTATCATCAGGAGACCAAAAGAAAT 899
 Db 860 CAATATTGATGATGCTCACCAGTGGCCAGCTTATGTTATCATCAGGAGACCAAAAGAAAT 801

QY 900 AATATCTTATGATGATGTTAAAGAGTCTATTGGGTAAAGTATTAAACAAAGTGGTATTGA 959
 Db 800 AATATCTTATGATGATGTTAAAGAGTCTATTGGGTAAAGTATTAAACAAAGTGGTATTGA 741

QY 960 GGTGAGGCTTAATCTCAGAAAGGCGCTGTTTCAAGGACGACGCTGTTCTTTTAAACTGGA 1019
 Db 740 GGTGAGGCTTAATCTCAGAAAGGCGCTGTTTCAAGGACGACGCTGTTCTTTTAAACTGGA 681

QY 1020 TTGCAAAATGGAAGTCAAGTGTATCAGACAACTAGCAGTTGTAGTGAATGGGAAAGA 1079
 Db 680 TTGCAAAATGGAAGTCAAGTGTATCAGACAACTAGCAGTTGTAGTGAATGGGAAAGA 621

QY 1080 AATGGGCTGCTGCTGGAAGAGTGGCATATCTTGGAAATGAACTCTCTGATGCAAGAGTG 1139
 Db 620 AATGGGCTGCTGCTGGAAGAGTGGCATATCTTGGAAATGAACTCTCTGATGCAAGAGTG 561

QY 1140 CTTGAAGAGAGTGGGCTTAAGTGGGCTCCTGCTGATGCTGTTCTTACGCCCGAGAGGC 1199
 Db 560 CTTGAAGAGAGTGGGCTTAAGTGGGCTCCTGCTGATGCTGTTCTTACGCCCGAGAGGC 501

QY 1200 TGTGGATACATTTGCAAAATGTAATGTTGGTGGCGTGGTG-CCATCCGAGAAATTGCGAGC 1258
 Db 500 TGTGGATACATTTGCAAAATGTAATGTTGGTGGCGTGGTG-CCATCCGAGAAATTGCGAGC 441

QY 1259 ACATTG-CCTACTAATGGAAGATT 1284
 Db 440 ACATTG-CCTACTAATGGAAGATT 414

RESULT 12
 ADB83231

ID ADB83231 standard; cDNA; 738 BP.

XX ADB83231;

XX 04-DEC-2003 (first entry)

XX Human cDNA sequence useful for the treatment of cancer (SeqID 1444).

DE human; prostate; cancer; cytostatic; gene therapy; vaccine;

KW immune response; gene; ss.

XX Homo sapiens.

XX WO2003050236-A2.

XX 19-JUN-2003.

XX 04-SEP-2002; 2002WO-US028214.

XX 07-DEC-2001; 2001US-00012697.

XX (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;

PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leskowitz D, Kita D;

XX Garcia V, Jones LM, Stache-Crain B, Scott EM;

XX WFI; 2003-513972/48.

XX New polynucleotides derived from human prostate, useful for modulating
 immune response to prevent or treat cancer.

PS Claim 1; SEQ ID NO 1444; 188pp; English.

XX This invention relates to novel isolated polynucleotides of human origin, particularly isolated from the human prostate. Specifically, it refers to the diagnostics and therapeutics comprising these novel human polynucleotides, and includes the derived probes, antisense oligonucleotides and antibodies thereof. The identification of these human prostate genes that can inhibit tumour growth is useful for understanding the progression and nature of complex diseases such as cancer, and hence they are important in the drug discovery process. The present invention describes these polynucleotides and encoded polypeptides as exhibiting cytostatic activity, and through gene therapy and/or vaccines they can be used to modulate the immune response for the prevention or treatment of cancers, particularly of the prostate, but also for breast, lung and colon cancer. This polynucleotide sequence is a human cDNA sequence useful for the treatment of cancer, used in an exemplification of the invention. NOTE: These sequences are not given in the specification but are provided on the WIPO website.

SQ Sequence 738 BP; 247 A; 97 C; 179 G; 215 T; 0 U; 0 Other;

Query Match 52.5%; Score 686.8; DB 8; Length 738;
Best Local Similarity 97.3%; Pred. No. 3.2e-184;
Matches 716; Conservative 3; Mismatches 15; Indels 2; Gaps 2;

QY 405 GGGTGCATTTGAGGAAATATCAAGCTACTCTTCYCATGTTTACATCTACTGATCTCA 464
DB 3 GGTGACATTTAGGAATATTCAGCTACTCTCCCATGTTTACATCTACTGATCTCA 62

QY 465 AAAAGTTCGAAATGATTCGAGNAGAGATATGATCTGKTTCTCTGTGTGAGAG 524
DB 63 AAAAGTTCGAAATGATTCGAGNAGAGATATGATCTGTTCTCTGTGTGAGAG 122

QY 525 CCATCAGTTTCGATGGAGTGAATTCAGAAAGAGTTTCGTGAAGTACCGAACCTCTGAA 584
DB 123 CCATCAGTTTCGATGGAGTGAATTCAGAAAGAGTTTCGTGAAGTACCGAACCTCTGAA 182

QY 585 TTTAAATCCAGCTAAACGGCTCTGTGACAGAGCTGGGATGGAGAAATATATGAAATGG 644
DB 183 TTTAAATCCAGCTAAACGGCTCTGTGACAGAGCTGGGATGGAGAAATATATGAAATGG 242

QY 645 CTGATTTATTTGCTAAAAGACATTTGATAGAGATGGTTACTTCAGGGTGG-AAAAT 703
DB 243 CTGATTTATTTGCTAAAAGACATTTGATAGAGATGGTTACTTCAGGGTGG-AAAAT 302

QY 704 GGCATCTACGAAATCGAGCTGCAACATATGTTGGATATAGATGTGGATTTGATTGGC 763
DB 303 GGCATCTACGAAATCGAGCT-GAACATATGTTGGATATAGATGTGGATTTGATTGGC 361

QY 764 CTATTGAGAGCAAGAGTATTAAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAA 823
DB 362 CTATTGAGAGCAAGAGTATTAAGATATGCTATTTTGGCAAGAGAGCTTAAGGAAA 421

QY 824 TAAACCTTTGGTTGCAATATTTGATGATGCTCTCACCAGTGGCCACATTTATGATCAG 883
DB 422 TAAACCTTTGGTTGCAATATTTGATGATGCTCTCACCAGTGGCCACATTTATGATCAG 481

QY 884 GAGACCAAAAGAAATATATCTTATGATGAAAGATGCTATTGGCATAGTTATTAA 943
DB 482 GAGACCAAAAGAAATATATCTTATGATGAAAGATGCTATTGGCATAGTTATTAA 541

QY 944 AGAAAGTGTATTGAGGTGAGGCTAATCTCAGAAAGGGCTGTTCAGAGCAGCGTGT 1003
DB 542 AGAAAGTGTATTGAGGTGAGGCTAATCTCAGAAAGGGCTGTTCAGAGCAGCGTGT 601

QY 1004 CTTCCTTTAAACTGGATGCAAAATGGAGTCACTGATCAGACAGCTAGCTGTAG 1063
DB 602 CTTCCTTTAAACTGGATGCAAAATGGAGTCACTGATCAGACAGCTAGCTGTAG 661

QY 1064 ATGAATCGAGAAAGAAATGGGCTGTGTGCGAAGAGAGTGGCATATCTTGGAAATGAAG 1123
DB 662 ATGAATCGAGAAAGAAATGGGCTGTGTGCGAAGAGAGTGGCATATCTTGTATTATT 721

QY 1124 TGTCTGATGAAGAGTG 1139
DB 722 TACTGTAGGAAATG 737

RESULT 13
ADB83156
ID ADB83156 standard; cDNA; 789 BP.
XX
AC ADB83156;
XX
XX 04-DEC-2003 (first entry)
XX Human cDNA sequence useful for the treatment of cancer (SeqID 1363).
XX human; prostate; cancer; cytostatic; gene therapy; vaccine;
XX immune response; gene; ss.
XX Homo sapiens.
OS
XX
PN WO2003050236-A2.
XX
PD 19-JUN-2003.
XX
PF 04-SEP-2002; 2002WO-US028214.
XX
PR 07-DEC-2001; 2001US-00012697.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Kassam A, Jamson G, Drmanac R;
PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;
PI Garcia V, Jones LW, Stache-Crain B, Scott EM;
XX
DR WPI; 2003-513972/48.
XX

PT New polynucleotides derived from human prostate, useful for modulating
PT immune response to prevent or treat cancer.
XX
XX Claim 1; SEQ ID NO 1369; 188pp; English.

XX This invention relates to novel isolated polynucleotides of human origin, particularly isolated from the human prostate. Specifically, it refers to the diagnostics and therapeutics comprising these novel human polynucleotides, and includes the derived probes, antisense oligonucleotides and antibodies thereof. The identification of these human prostate genes that can inhibit tumour growth is useful for understanding the progression and nature of complex diseases such as cancer, and hence they are important in the drug discovery process. The present invention describes these polynucleotides and encoded polypeptides as exhibiting cytostatic activity, and through gene therapy and/or vaccines they can be used to modulate the immune response for the prevention or treatment of cancers, particularly of the prostate, but also for breast, lung and colon cancer. This polynucleotide sequence is a human cDNA sequence useful for the treatment of cancer, used in an exemplification of the invention. NOTE: These sequences are not given in the specification but are provided on the WIPO website.

XX
SQ Sequence 789 BP; 235 A; 131 C; 199 G; 216 T; 0 U; 8 Other;

Query Match 49.6%; Score 647.6; DB 8; Length 789;
Best Local Similarity 94.8%; Pred. No. 4.5e-173;
Matches 729; Conservative 5; Mismatches 22; Indels 13; Gaps 6;

QY 150 GGGCCGGGGAGGAGCAGCAAGCATCCCTTCAGAGACATTAAAGCCTCGGGGGTCCC 209
DB 14 GGGCCGGGGAGGAGCAGCAAGCATCCCTTCAGAGACATTAAAGCCTCGGGGGTCCC 73

QY 210 GCTCATTTGGTGGTCTCTGGTGGGCTTCAGGGCCCTTCAGAGCTGTATGGGT 269
DB 74 GCTCATTTGGTGGTCTCTGGTGGGCTTCAGGGCCCTTCAGAGCTGTATGGGT 133

QY 270 TTGGACAGACCATGATGAAATTTGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 329
 DB 134 TTGACAGACCATGATGAAATTTGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 193
 QY 330 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 389
 DB 194 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 253
 QY 390 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 449
 DB 254 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 313
 QY 450 TCCACTGATCTTCAAAAGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTG 509
 DB 314 TCCACTGATCTTCAAAAGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTG 373
 QY 510 CTCTGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 569
 DB 374 CTCCTGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 433
 QY 570 GACCGAAGCTCTGAAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 629
 DB 434 GACCGAAGCTCTGAAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 493
 QY 630 ATTATATGAAATGCTCATTTTATTTGCTTAAAGACATTTGATAGATGCTGTTACTT 689
 DB 494 ATTATATGAAATGCTCATTTTATTTGCTTAAAGACATTTGATAGATGCTGTTACTT 553
 QY 690 GCAGGTGG-AAAATGGCATCTACGAAATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 748
 DB 554 GCAGGTGG-AAAATGGCATCTACGAAATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 612
 QY 749 TGGATATTGATTTGGCTTATTGACAGCAAGAGTATTAGATATGCTATTTTTGGCAAG 808
 DB 613 TGGATATTGATTTGGCTTATTGACAGCAAGAGTATTAGATATGCTATTTTTGGCAAG 672
 QY 809 AGAGCTTAAGGAAATTAACCTTTTGGTTGCAATATTGATGATGCTCACCAGTGGCC 868
 DB 673 AGAGCTTAAGGAA--TAACTTTTGGTT--GCATATTGATGATGCTCACCAGTGGCC 728
 QY 869 ACATTATGATCATGAGCAACCAAAAGAAATAATATCTATGATCTAAA 917
 DB 729 A---TTTTGTHCAGGACCA--AGATATTATGATGATGATGATGATGATGATGATGATGAT 770

RESULT 14
 ABQ89138
 ID ABQ89138 standard; cDNA; 838 BP.
 AC ABQ89138;
 DT 27-SEP-2002 (first entry)
 DE Human prostate expressed polynucleotide SEQ ID NO 394.
 XX Human; prostate; cytostatic; tumour; cancer; vaccine; gene therapy; gene;
 KW ss.
 XX Homo sapiens.
 XX WO200255700-A2.
 XX 18-JUL-2002.
 XX 07-DEC-2001; 2001WC-US047349.
 XX 07-DEC-2000; 2000US-0254648P.
 XX 13-MAR-2001; 2001US-0275688P.
 XX (CHIR) CHIRN CORP.
 XX (HYSE-) HYSEQ INC.
 PI Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;

PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Ieshkowitz D, Kita D;
 PI Garcia V, Jones WL, Stache-Crain B, Scott EM;
 XX MPI; 2002-557824/59.
 XX New genes and gene products isolated from human prostate, useful for
 PT treating or diagnosing tumor or cancer (e.g. prostate cancer or breast
 PT cancer), or as vaccines for treating or preventing these diseases.
 XX Claim 1; SEQ ID NO 394; 186pp + Sequence Listing; English.
 XX The invention relates to an isolated polynucleotide comprising any of
 CC 1477 sequences or its fragment, degenerate variant, antisense or
 CC complement. The polynucleotides and gene products are useful for treating
 CC or diagnosing tumor or cancer (e.g. prostate cancer, breast cancer, lung
 CC cancer or medullary carcinoma) in a subject (e.g. cattle, dogs, cats,
 CC rabbits, horse or human). The polynucleotides and polypeptides are also
 CC useful as vaccines for treating or preventing these diseases. The
 CC polynucleotides are useful for gene therapy. The present sequence is that
 CC of one of a group of polynucleotides (ABQ8745-ABQ9015) disclosed
 CC electronically as sequences of the invention. However only 1271
 CC polynucleotide sequences are given, whereas 1477 polynucleotides and 91
 CC proteins are claimed. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pot_sequence
 XX Sequence 838 BP; 241 A; 151 C; 199 G; 227 T; 0 U; 20 Other;
 QY Query Match 48.8%; Score 636.6; DB 6; Length 838;
 DB Best Local Similarity 94.7%; Pred. No. 6.2e-170; Indels 10; Gaps 6;
 DB Matches 715; Conservative 4; Mismatches 26; Indels 10; Gaps 6;
 QY 150 GCGCCGGGAGGACGACAAAGGATCCCTCTGAAGAACATTAAAGCAGCTCGCGGGGTCCC 209
 DB 71 GCGCCGGGAGGACGACAAAGGATCCCTCTGAAGAACATTAAAGCAGCTCGCGGGGTCCC 130
 QY 210 GCTCATTTGGTGGTTCCTGCGTGGCGCTGATTCAGGGGCTTCCAGAGTGTATGGGT 269
 DB 131 GCTCATTTGGTGGTTCCTGCGTGGCGCTGATTCAGGGGCTTCCAGAGTGTATGGGT 190
 QY 270 TTCGACAGACCATGATGAAATTTGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 329
 DB 191 TTCGACAGACCATGATGAAATTTGAGAACTGTGCGCAAAACAATTTGGTGGCAAGTTCATCG 250
 QY 330 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 389
 DB 251 AAGAGTTCTGAGTTTCAAAAGACAGCTCTACCTCAGTACAGTGCATCATAGATTTCT 310
 QY 390 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 449
 DB 311 TAAATATATATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 370
 QY 450 TCCACTGATCTTCAAAAGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTG 509
 DB 371 TCCACTGATCTTCAAAAGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTGCAAGATGTTG 430
 QY 510 CTCTGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 569
 DB 431 CTCCTGTTGTGAGCGCCATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTGCTGAGT 490
 QY 570 GACCGAAGCTCTGAAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 629
 DB 491 GACCGAAGCTCTGAAATTTAAATCCAGCTTAAACGGCTCTGTCGACAAAGCTGGATGGAGA 550
 QY 630 ATTATATGAAATGCTCATTTTATTTGCTTAAAGACATTTGATAGATGCTGTTACTT 689
 DB 551 ATTATATGAAATGCTCATTTTATTTGCTTAAAGACATTTGATAGATGCTGTTACTT 610
 QY 690 GCAGGTGG-AAAATGGCATCTACGAAATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 748
 DB 611 GCAGGTGG-AAAATGGCATCTACGAAATGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 669
 QY 749 TGGATATTGATTTGGCTTATTGACAGCAAGAGTATTAGATATGCTATTTTTGGCAAG 808

Db 670 TGGATATTGATGGCTATGAGCAAGAGTATTAAATATGCTTTTGG-TAAG 728
 QY 809 AGAAGCTTAAGAAATAAACTTTTGGTTTGAATATGATGATGCTCTACCAATGGCC 868
 Db 729 ANAAGCTTAAG---AATAAACTTTTGGTTT--GCATATGATGATGCTCTACCA--TGGC 781
 QY 869 ACATTTATGATCAGGAGACCAAAAGAAATATA 903
 Db 782 CCNTTNTGNTCAGGAGACCAAAAAA 816

RESULT 15
 ADB82082
 ID ADB82082 standard; cDNA; 745 BP.
 AC ADB82082;
 XX
 DT 04-DEC-2003 (first entry)
 DE Human cDNA sequence useful for the treatment of cancer (SeqID 394).
 KW humar; prostate; cancer; cytostatic; gene therapy; vaccine;
 KW immune response; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WC0203050236-A2.
 XX
 PD 19-JUN-2003.
 XX
 PF 04-SEP-2002; 2002WO-US020214.
 XX
 PR 07-DEC-2001; 2001US-00012697.
 XX
 PA (CHIR } HYSEQ INC.
 PA (HYSE-) CHIRON CORP.

EScobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;
 FI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;
 PI Garcia V, Jones LW, Stache-Crain B, Scott EM;
 XX WPI; 2003-513972/48.
 DR
 XX New polynucleotides derived from human prostate, useful for modulating
 PT immune response to prevent or treat cancer.
 PT
 XX Claim 1; SEQ ID NO 394; 188pp; English.

This invention relates to novel isolated polynucleotides of human origin,
 particularly isolated from the human prostate. Specifically, it refers to
 the diagnostics and therapeutics comprising these novel human
 polynucleotides, and includes the derived probes, antisense
 oligonucleotides and antibodies thereof. The identification of these
 human prostate genes that can inhibit tumour growth is useful for
 understanding the progression and nature of complex diseases such as
 cancer, and hence they are important in the drug discovery process. The
 present invention describes these polynucleotides and encoded
 polypeptides as exhibiting cytostatic activity, and through gene therapy
 and/or vaccines they can be used to modulate the immune response for the
 prevention or treatment of cancers, particularly of the prostate, but
 also for breast, lung and colon cancer. This polynucleotide sequence is a
 human cDNA sequence useful for the treatment of cancer, used in an
 exemplification of the invention. NOTE: These sequences are not given in
 the specification but are provided on the WIPO website.

Sequence 745 BP; 216 A; 129 C; 185 G; 204 T; 0 U; 11 Other;
 Query Match 48.1%; Score 627.4; DB 8; Length 745;
 Best Local Similarity 95.0%; Pred. No. 2.4e-167;
 Matches 704; Conservative 4; Mismatches 23; Indels 10; Gaps 6;
 QY 150 GGCCCGGGGAGGAGCAAGGATCCCTCGAAGACATTAAGCACTGGCGGGGTCCC 209

Db 14 GGCCCGGGGAGGAGCAAGGATCCCTCGAAGACATTAAGCACTGGCGGGGTCCC 73
 QY 210 GCTCATTTGGCTGGGCTCTCGCTGCGGCCCTTGGATTTCAGGGGCTTCCAGAGTGTATGGGT 269
 Db 74 GCTCATTTGGCTGGGCTCTCGCTGCGGCCCTTGGATTTCAGGGGCTTCCAGAGTGTATGGGT 133
 QY 270 TTGACAGACCATGATGAAATTTGAGAAATGTCGCAACCAATTTGGTGCACAAATTCATCG 329
 Db 134 TTGACAGACCATGATGAAATTTGAGAAATGTCGCAACCAATTTGGTGCACAAATTCATCG 193
 QY 330 AAGAAGTCTCTGAAGTTTCAAAAGACAGCTCTACTCTCAGTAGATGCCCATATAGAAATTTCT 389
 Db 194 AAGAAGTCTCTGAAGTTTCAAAAGACAGCTCTACTCTCAGTAGATGCCCATATAGAAATTTCT 253
 QY 390 TAATTATATATAGAGGTGACATTGTAGAAATATCAAGCTACTTCTTCATGTTTACA 449
 Db 254 TAATTATATATAGAGGTGACATTGTAGAAATATCAAGCTACTTCTTCATGTTTACA 313
 QY 450 CCTACTGATCTTCAAAAGTTTCAAAAGTTTCAAAAGTTTCAAAAGTTTCAAAAGTTTCAAAAGTTT 509
 Db 314 CCTACTGATCTTCAAAAGTTTCAAAAGTTTCAAAAGTTTCAAAAGTTTCAAAAGTTTCAAAAGTTT 373
 QY 510 CTCTGTTGTGAGACGCGCATCAGTTTCGATGAGTGAATTCAGAAAGGAGTTCGTGAAGT 569
 Db 374 CTCTGTTGTGAGACGCGCATCAGTTTCGATGAGTGAATTCAGAAAGGAGTTCGTGAAGT 433
 QY 570 GACCGAACCTCTGAATTTAAATCCAGCTAAAAGGCTCTGCAAGACATTTGATAGAGATGGTTACTT 629
 Db 434 GACCGAACCTCTGAATTTAAATCCAGCTAAAAGGCTCTGCAAGACATTTGATAGAGATGGTTACTT 493
 QY 630 ATTATATGAATATGCTCATTTTATTTTCTTAAAGACATTTGATAGAGATGGTTACTT 689
 Db 494 ATTATATGAATATGCTCATTTTATTTTCTTAAAGACATTTGATAGAGATGGTTACTT 553
 QY 690 GCAGGCTGG-AAAATGGCATACTACGAAATGCGAGCTGGAAACATAGTGTGGATATAGATG 748
 Db 554 GCAGGCTGGAAATGGCATACTACGAAATGCGAGCT-GAACATAGTGTGGATATAGATG 612
 QY 749 TGGATATTGATTTGGCTTATGACAGCAAGAGTATTAGATATGGCTATTTTGGCAAG 808
 Db 613 TGGATATTGATTTGGCTTATGACAGCAAGAGTATTAGATATGGCTATTTTTGG-TAAG 671
 QY 809 AGAAGCTTAAGGAATAAACTTTTGGTTTGGCAATATTGATGATGCTCTCAACCAATGGCC 868
 Db 672 ANAAGCTTAAG--AATAAACTTTTGGTTT--GCATATTGATGATGCTCTCAACCA--TGGC 724
 QY 869 ACATTTATGATCAGGAGACC 889
 Db 725 CCNTTNTGNTCAGGAGACC 745

Search completed: May 10, 2004, 08:07:58
 Job time : 570.652 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 ; Search time 3773.33 Seconds
(without alignments)
10327.779 Million cell updates/sec

Title: US-09-930-440b-3

Perfect score: 1305

Sequence: 1 atggactcggggagaagg.....ataattcgtccaaaataag 1305

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estha.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_estl.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1011.2	77.5	1201	9	AL530667
2	927.8	71.1	1201	9	AL542712
3	978	67.3	1088	9	AL533530
4	869.8	66.7	1201	9	AL537974

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	853	65.4	1201	9	AL557407
6	849.8	65.1	1038	9	AL540564
7	830.4	63.6	1201	13	BM406343
8	821.2	62.9	990	12	BM457667
9	801.4	61.4	904	12	BM457667
10	798.6	61.2	881	13	BQ226032
11	786.8	60.3	1201	9	AL538941
12	768.4	58.9	1045	29	AY412329
13	751.6	57.6	982	12	BQ674331
14	745.2	57.1	1201	13	BX343798
15	737.8	56.5	874	13	BQ438805
16	727	55.7	1201	9	AL530666
17	722.4	55.4	834	9	AL559436
18	718.4	55.0	805	9	AUL26230
19	718.4	55.0	883	9	AL560379
20	712.4	54.6	823	12	BI837810
21	709.8	54.4	804	12	BI92760
22	708.4	54.3	872	12	BI769833
23	707	54.2	784	12	BI953110
24	706	54.1	884	13	BQ215134
25	705.2	54.0	1036	9	AL545855
26	705	54.0	1087	13	BQ67421
27	703.4	53.9	753	12	BI717075
28	701.4	53.7	802	12	BI97315
29	700.4	53.7	764	14	CA772419
30	699.8	53.6	807	12	BI97820
31	691.6	53.0	912	12	BI462549
32	657	50.3	1201	9	AL569567
33	656	50.3	1201	13	BX463155
34	650	49.8	905	12	BG682092
35	642.6	49.2	732	12	BI600714
36	641	49.1	1201	9	AL537973
37	633.8	48.6	664	12	BM792481
38	633.8	48.6	859	14	CD358582
39	630.8	48.3	1201	13	BX406342
40	624.4	47.8	1045	29	AY412331
41	622	47.7	732	10	BZ253152
42	610.6	46.8	651	10	BZ253161
43	609.8	46.7	636	14	CD672433
44	609.4	46.7	879	14	CD511863
45	609	46.7	1044	9	AL565152

ALIGNMENTS

RESULT 1

AL530667

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL530667 1201 bp mRNA linear EST 23-MAY-2003
AL530667 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
CDNA clone CS0DD008YG12 5-PRIME, mRNA sequence.

AL530667

AL530667.2 GI:31068500

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1201)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 13, 2001 this sequence version replaced gi:12794160.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91036 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

6598.f For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DD008Y06021&cluster=6598.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :


```

115 ATGGACTCGTGGAGAGAGGGGGCCGCCACCTCCGCTCTCCAAACCGCGGGGGCGACCGTCC 174
116 |||||
117 |||||
118 |||||
119 |||||
120 CGGSGCCGGCCGCCAGAGCTCAGCGCAACTCTCGCGCGGCCCGCAGGGCGGAGGTGTGGAG 120
121 |||||
122 |||||
123 |||||
124 |||||
125 CGGSGCCGGCCGCCAGAGCTCAGCGCAACTCTCGCGCGGCCCGCAGGGCGGAGGTGTGGAG 234
126 |||||
127 |||||
128 |||||
129 |||||
130 AAGCCCCCGCACCTCGGCGAGCCCTAATCTGCGCGCGGGGAGCAGCAAAAGGATCCCCCTG 180
131 |||||
132 |||||
133 |||||
134 AAGCCCCCGCACCTCGGCGAGCCCTAATCTGCGCGCGGGGAGCAGCAAAAGGATCCCCCTG 294
135 |||||
136 |||||
137 |||||
138 |||||
139 AAGAACATTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCA 240
140 |||||
141 |||||
142 |||||
143 |||||
144 AAGAACATTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCA 354
145 |||||
146 |||||
147 |||||
148 |||||
149 GATTTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAAATTCAGAAATG 300
150 |||||
151 |||||
152 |||||
153 |||||
154 GATTTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAAATTCAGAAATG 414
155 |||||
156 |||||
157 |||||
158 |||||
159 GCCAAACAATTTGGTGCAAGTTCATCGAAGAGTTCGAAAGTTCGAAAGTTCGAAAGTTC 360
160 |||||
161 |||||
162 |||||
163 |||||
164 GCCAAACAATTTGGTGCAAGTTCATCGAAGAGTTCGAAAGTTCGAAAGTTCGAAAGTTC 474
165 |||||
166 |||||
167 |||||
168 |||||
169 ACCTCAGTAGATGCATCATAGAAATTTCTTAATATATATATATATATATATATATATAT 420
170 |||||
171 |||||
172 |||||
173 |||||
174 ACCTCAGTAGATGCATCATAGAAATTTCTTAATATATATATATATATATATATATATAT 534
175 |||||
176 |||||
177 |||||
178 |||||
179 AATATTCAAGCTCTCTCCAGATTTTACATCCTACTGATCTTCAAAAGTTCGAAATG 594
180 |||||
181 |||||
182 |||||
183 |||||
184 ATTCGAGAGAGAGATATGATCTGKTTTCTGTTCTGTTGAGAGCGCATCAGTTTCGATGG 540
185 |||||
186 |||||
187 |||||
188 |||||
189 ATTCGAGAGAGAGATATGATCTGKTTTCTGTTCTGTTGAGAGCGCATCAGTTTCGATGG 654
190 |||||
191 |||||
192 |||||
193 |||||
194 AGTCAAAATTCGAAAGAGGTTCTGAGTGCAGGAGCCTCTGAAATTTAAATCCAGCTAAA 600
195 |||||
196 |||||
197 |||||
198 |||||
199 AGTCAAAATTCGAAAGAGGTTCTGAGTGCAGGAGCCTCTGAAATTTAAATCCAGCTAAA 714
200 |||||
201 |||||
202 |||||
203 |||||
204 CGGCTCGTGCAGCAAGAGTGGGATGGAATTTATATGAAATGCTCATTTTATTTGCT 660
205 |||||
206 |||||
207 |||||
208 |||||
209 CGGCTCGTGCAGCAAGAGTGGGATGGAATTTATATGAAATGCTCATTTTATTTGCT 774
210 |||||
211 |||||
212 |||||
213 |||||
214 AAAAGACATTTGATAGAGATGGGTACTTGCAGGGTGGAAATGGCATACTATGAAATG 834
215 |||||
216 |||||
217 |||||
218 |||||
219 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATGATTTGCTATTTGAGAGCAAG 779
220 |||||
221 |||||
222 |||||
223 |||||
224 CGAGCTGGAACATAGTGTGGATATAGATGTGGATATGATTTGCTATTTGAGAGCAAG 893
225 |||||
226 |||||
227 |||||
228 |||||
229 AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAAGAAATTAAGAAATTTTGGTTG 839
230 |||||
231 |||||
232 |||||
233 |||||
234 AGTATTAGATATGGCTATTTTGGCAAGAGAGCTTAAAGAAATTAAGAAATTTTGGTTG 952
235 |||||
236 |||||
237 |||||
238 |||||
239 CAATATTGATGGATCTCCACCAATGGCCACA-TTATGATATCAGAGACCAAAAGAA 898
240 |||||
241 |||||
242 |||||
243 |||||
244 CAATATTGATGGATCTCCACCAATGGCCACA-TTATGATATCAGAGACCAAAAGAA 1012
245 |||||
246 |||||
247 |||||
248 |||||
249 TAATATCTTATGATTAAGATGCTATTGGGATAAGTTTATTAAGAAAGAGTGTATTG 958
250 |||||
251 |||||
252 |||||
253 |||||
254 TAATATCTTATGATTAAGATGCTATTGGGATAAGTTTATTAAGAAAGAGTGTATTG 1072
255 |||||
256 |||||
257 |||||
258 |||||
259 AGGTGAGGCTAATCTCGAAAGAGGCTGTTCGAAAGCAGAGCTGTCTTCTTTAAAGCTGG 1018
260 |||||
261 |||||
262 |||||
263 |||||
264 AGGTGAGGCTAATCTCGAAAGAGGCTGTTCGAAAGCAGAGCTGTCTTCTTTAAAGCTGG 1129
265 |||||
266 |||||
267 |||||
268 |||||
269 ATTGCAAAATGGAGTCACTGATATCAGACAGAGTATGATGATGATGATGATGATGATG 1075
270 |||||
271 |||||
272 |||||
273 |||||
274 ATTGCAAAATGGAGTCACTGATATCAGACAGAGTATGATGATGATGATGATGATGATG 1186
275 |||||
276 |||||
277 |||||
278 |||||

```

```

AL533530 1533530 1988 bp mRNA linear EST 31-MAY-2003
LOCUS AL533530 Homo sapiens ADULT BRAIN Homo sapiens cDNA clone
DEFINITION CSODN004YA17 5-PRIME, mRNA sequence.
ACCESSION AL533530
VERSION AL533530.2 GI:31260610
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1088)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLES Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12737023.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6598.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODN004YA09QP1&cluster=6598.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/InvitroGer Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODN004YA09QP1.
FEATURES
source
1..1088
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODN004YA17"
/tissue_type="ADULT BRAIN"
/dev_stage="adult"
/clone_lib="Homo sapiens ADULT BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
ORIGIN
Query Match 67.3%; Score 878; DB 9; Length 1088;
Best Local Similarity 97.6%; Pred. No. 8.3e-220;
Matches 943; Conservative 10; Mismatches 5; Indels 8; Gaps 6;
QY 1 ATGGACTCGTGGAGAGAGGGGGCCGCCACCTCCGCTCTCCAAACCGCGGGGGCGACCGTCC 60
DB 79 ATGGACTCGTGGAGAGAGGGGGCCGCCACCTCCGCTCTCCAAACCGCGGGGGCGACCGTCC 138
QY 61 CGGSGCCGGCCGCCAGAGCTCAGCGCAACTCTCGCGCGGCCCGCAGGGCGGAGGTGTGGAG 120
DB 139 CGGSGCCGGCCGCCAGAGCTCAGCGCAACTCTCGCGCGGCCCGCAGGGCGGAGGTGTGGAG 198
QY 121 AAGCCCCCGCACCTCGGCGAGCCCTAATCTGCGCGCGGGGAGGAGCAAAAGGATCCCCCTG 180
DB 199 AAGCCCCCGCACCTCGGCGAGCCCTAATCTGCGCGCGGGGAGGAGCAAAAGGATCCCCCTG 258
QY 181 AAGAACATTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCA 240
DB 259 AAGAACATTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCACTTAAAGCA 318
QY 241 GATTTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAAATTCAGAAATG 300
DB 319 GATTTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAAATTCAGAAATG 378
QY 301 GCCAAACAATTTGGTGCAAGTTCATCGAAGAGTTCGAAAGTTCGAAAGTTCGAAAGTTC 360
DB 379 GCCAAACAATTTGGTGCAAGTTCATCGAAGAGTTCGAAAGTTCGAAAGTTCGAAAGTTC 438
QY 361 ACCTCAGTAGATGCATCATAGAAATTTCTTAATATATATATATATATATATATATATAT 420

```



```

COMMENT      On Feb 15, 2001 this sequence version replaced gi:12370832.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6598.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DE002DA01QP1&cluster=6598.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DE002DA01QP1.

FEATURES
source
1..1038
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE002YB02"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match      65.1%; Score 849.8; DB 9; Length 1038;
Best Local Similarity 96.1%; Pred. No. 2.2e-212;
Matches 889; Conservative 20; Mismatches 12; Indels 4; Gaps 4;

QY 1 ATGGACTCGGTGGAGAGGGGGCGCCACCTCGCTCCACCCGCGGGGGCCACGCTCC 60
DB 115 ATGGACTCGGTGGAGAGGGGGCGCCACCTCGCTCCACCCGCGGGGGCCACGCTCC 174
QY 61 CGGGGGCGCGCGCGGAGCTGCAGCGCACTCTCGCGGGGGCGGAGCGGAGTGGAG 120
DB 175 CGGGGGCGCGCGCGGAGCTGCAGCGCACTCTCGCGGGGGCGGAGCGGAGTGGAG 234
QY 121 AAGCCCCCGACCTGCGAGCCCTTAATCTTGGCCCGGGGAGCGGAGGAGCATCCCGCTG 180
DB 235 AAGCCCCCGACCTGCGAGCCCTTAATCTTGGCCCGGGGAGCGGAGGAGCATCCCGCTG 294
QY 181 AAGACATTAAAGCACTCGCGGGGGTCCCGCTCATTTGGCTGGGTCTGCTGGGGCCCTG 240
DB 295 AAGACATTAAAGCACTCGCGGGGGTCCCGCTCATTTGGCTGGGTCTGCTGGGGCCCTG 354
QY 241 GATTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCGAGATGTG 300
DB 355 GATTCAGGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTCGAGATGTG 414
QY 301 GCGCAACAATTTGGTGCACAAGTTCTGAGAGAGTTCGAGTTTCAAAAGCAGCTCT 360
DB 415 GCGCAACAATTTGGTGCACAAGTTCTGAGAGAGTTCGAGTTTCAAAAGCAGCTCT 474
QY 361 ACCTCACTAGTCCCATCATAGAAATTCCTAATTATYATATCAGTGTGACATGTGAGGA 420
DB 475 ACCTCACTAGTCCCATCATAGAAATTCCTAATTATYATATCAGTGTGACATGTGAGGA 534
QY 421 AATATTCAAGTACTTCTTCATGTTTACATCTTACTGATCTTCAAAAAGTTCGAGAAATG 480
DB 535 AATATTCAAGTACTTCTTCATGTTTACATCTTACTGATCTTCAAAAAGTTCGAGAAATG 594
QY 481 ATTTCGAGAGAGGATATGATTCGTTCTTCTGTTGAGAGCGCATCATGTTTCGATGG 540
DB 595 ATTTCGAGAGAGGATATGATTCGTTCTTCTGTTGAGAGCGCATCATGTTTCGATGG 654
QY 541 ACTGGAATTCAGAAAGGATTCGTTGAGTGAAGTACCGGACCTCTGAAATTAATTCAGCTAA 600
DB 655 AGTGAATTCAGAAAGGATTCGTTGAGTGAAGTACCGGACCTCTGAAATTAATTCAGCTAA 714
QY 601 CGGCTCTCGTCCAGACACTGGGATGGAGAAATATATGAAATGGCTCAATTTATTTTGTCT 660

```

```

DB 715 CGGCCCTCGTGCACAGACTGGGATGGAGATTAATGAAATGGCTCATTTATTTTGTCT 774
QY 661 AAAAGACATTTGATAGAGATGGGTTACTTGCAGGGTGG-AAAATGGCATCTACGAAATG 719
DB 775 AAAAGACATTTGATAGAGATGGGTTACTTGCAGGGTGG-AAAATGGCATCTACGAAATG 834
QY 720 CGAGCTGGAGACATAGTGTGATATAGATATGATTTGGTATATGATTTGCATTCAGAGCAAG 779
DB 835 CGAGCT-GAACAATAGTGTGATATAGATATGATTTGGTATATGATTTGCATTCAGAGCAAG 893
QY 780 AGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGCAATATAAAC-TTTTGGTTT 838
DB 894 AGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGCAATATAAAC-TTTTGGTTT 953
QY 839 GCAATATTGATGATGTCTCACCAATGGCCACAT-TTATGATCAGAGAGACCAAAAGAA 897
DB 954 GCRATATTGATGATGTCTCTCMCATAGGCMCATATTTTGTTCAGGGGGCCCAAAAAA 1013
QY 898 ATAATATCTTATGATGTAAAGATG 922
DB 1014 AAAAAGWCTTTTGTGCTGATAAAAG 1038

RESULT 7
LOCUS      BX406343      1201 bp      mRNA      linear      EST 15-MAY-2003
DEFINITION BX406343 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
ACCESSION . CS0DM011YG12 5-PRIME, mRNA sequence.
VERSION     BX406343
KEYWORDS    BX406343.1 GI:30764679
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE    1 (bases 1 to 1201)
AUTHORS     Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6598.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AM011BD06QP1&cluster=6598.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0AM011BD06QP1.

FEATURES
source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM011YG12"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match      63.6%; Score 830.4; DB 13; Length 1201;
Best Local Similarity 93.3%; Pred. No. 3e-207;
Matches 940; Conservative 21; Mismatches 33; Indels 13; Gaps 9;

QY 1 ATGGACTCGGTGGAGAGGGGGCGGCGCCACCTCGCTCTCCAAACCCGCGGGGGCCACGCTCC 60

```



```

Db 160 ATGACCTCGGTGAGAAAGGGGGCGGCACCTCGTCTCCAAACCGCGGGGCGACCGCTCC 219
Qy 61 CGGGGGCGGGCGGCGAAGCTGACGCAAACTCTCGCGCGGGCCAGGCGCCGAGGCTGTGGAG 120
Db 220 CGGGGGCGGGCGGCGAAGCTGACGCAAACTCTCGCGCGGGCCAGGCGCCGAGGCTGTGGAG 279
Qy 121 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGCGCGGGAGGAGCAGCAAAAGGATCCCCCTG 180
Db 280 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGCGCGGGAGGAGCAGCAAAAGGATCCCCCTG 339
Qy 181 AAGAACATTAAGACACTGGCGGGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 340 AAGAACATTAAGACACTGGCGGGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 399
Qy 241 GATTCAGGGGCTTCACAGAGTGTATGGTTCACAGACCATGATGATGATGATGATGATGATGATG 300
Db 400 GATTCAGGGGCTTCACAGAGTGTATGGTTCACAGACCATGATGATGATGATGATGATGATGATG 459
Qy 301 GCCAAACAATTTGGTGCACAGTTCATCGAAGAGTTCGAGTTTCAAAAGACAGCTCT 360
Db 460 GCCAAACAATTTGGTGCACAGTTCATCGAAGAGTTCGAGTTTCAAAAGACAGCTCT 519
Qy 361 ACCTCACTAGATGCATCATAGATTTCTTAATATATATATATATATATATATATATATATATAT 420
Db 520 ACCTCACTAGATGCATCATAGATTTCTTAATATATATATATATATATATATATATATATATAT 579
Qy 421 AATATTCAAGCTACTCTTCATGCTTTACATCTCTACTGATCTTCAAAAGTTTCAGAAATG 480
Db 580 AATATTCAAGCTACTCTTCATGCTTTACATCTCTACTGATCTTCAAAAGTTTCAGAAATG 639
Qy 481 ATTCGAGAAAGATGATGATTTCTGTTTCTGTTGAGAGCCATCATGTTTCGATGG 540
Db 640 ATTCGAGAAAGATGATGATTTCTGTTTCTGTTGAGAGCCATCATGTTTCGATGG 699
Qy 541 AGTGAATTCAGAAAGAGTTCGTTGAGTGCAGCAACTCTGATTTAAATCCAGCTAAA 600
Db 700 AGTGAATTCAGAAAGAGTTCGTTGAGTGCAGCAACTCTGATTTAAATCCAGCTAAA 759
Qy 601 CGGCCTCGTCGCAAGACTGGGATGAGAAATATATGAAATGGCTCTATTTATTTTGT 660
Db 760 CGGCCTCGTCGCAAGACTGGGATGAGAAATATATGAAATGGCTCTATTTATTTTGT 819
Qy 661 AAGACATTTGATAGAGTGGTTACTTCGAGGCTG-BAATGCGCATACTAGCAATG 719
Db 820 AAGACATTTGATAGAGTGGTTACTTCGAGGCTG-BAATGCGCATACTAGCAATG 879
Qy 720 CGAGCTGGAACATAGTGTGATATAGATG-GGATATTGATTGGCTATTTCAGAGCAAG 779
Db 880 CGAGC-GGAACATAGTGTGATATAGATG-GGATATTGATTGG-CTATTGAGAGC-AAG 936
Qy 780 AGTATTAGATATGGCTATTGTCAGAAAGAGCTTTAAGAAATAAAATCTTTGGTTG 839
Db 937 AGTATTAGATATGGCTATTGTCAGAAAGAG- -GCTTTAGAAATAACACTTTTGGTGT 994
Qy 840 CAATATTGATGATGCTCCACATGCGCATATTATGATCAGGACCAAAAGAT 899
Db 995 CNAATTGATGATGCTC-MCATGCGCATATTATGATCAGGAGCAAAA- -GAT 1051
Qy 900 AATATCTTATGATGATAAGATGCTATTGGGATAAGTTTATTAAGAAAGTGTATTGA 959
Db 1052 WATATCTTATGATGATAAGATGCTATTGGGATA- -GTTTATTAAGAAAGTGTATTGA 1109
Qy 960 GGTGAGCTAATCTCAGAAAGGGCTGTTCAGAGCAGACAGCTGTCTT 1006
Db 1110 G--WGAGCTAATVMTAAAGGGCTTTTCAAAGMRMGSYGTTTTTTT 1154

```

RESULT 8
 AL558536 990 bp mRNA linear EST 31-MAY-2003
 LOCUS AL558536 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
 DEFINITION Homo sapiens cDNA clone CS0DJ006YB11 5-PRIME, mRNA sequence.

ACCESSION AL558536 GI:31280334
 VERSION AL558536.2
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1. (bases 1 to 990)
 AUTHORS Li, W.B., Graber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12903150.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6598.f For more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DJ006CA06QPI&cluster=6598.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0DJ006CA06QPI.
 FEATURES
 Location/Qualifiers
 1..990
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ006YB11"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_line="JURKAT"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and Eco3 V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 62.9%; Score 821.2; DB 9; Length 990;
 Best Local Similarity 98.4%; Pred. No. 7.4e-205;
 Matches 872; Conservative 8; Mismatches 1; Indels 5; Gaps 5;
 Qy 1 ATGACCTCGGTGAGAAAGGGGGCGGCACCTCGTCTCCAAACCGCGGGGCGACCGCTCC 60
 Db 109 ATGACCTCGGTGAGAAAGGGGGCGGCACCTCGTCTCCAAACCGCGGGGCGACCGCTCC 168
 Qy 61 CGGGGGCGGGCGGCGAAGCTGACGCAAACTCTCGCGGGCCAGGCGCCGAGGCTGTGGAG 120
 Db 169 CGGGGGCGGGCGGCGAAGCTGACGCAAACTCTCGCGGGCCAGGCGCCGAGGCTGTGGAG 228
 Qy 121 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGCGCGGGAGGAGCAGCAAAAGGATCCCCCTG 180
 Db 229 AAGCCCCCGCACCTGGCAGCCCTAAATCTGTGCGCGGGAGGAGCAGCAAAAGGATCCCCCTG 288
 Qy 181 AAGAACATTAAGACACTGGCGGGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 Db 289 AAGAACATTAAGACACTGGCGGGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 348
 Qy 241 GATTCAGGGGCTTCACAGAGTGTATGGTTCACAGACCATGATGATGATGATGATGATGATG 300
 Db 349 GATTCAGGGGCTTCACAGAGTGTATGGTTCACAGACCATGATGATGATGATGATGATGATG 408
 Qy 301 GCCAAACAATTTGGTGCACAGTTCATCGAAGAGTTCGAGTTTCAAAAGACAGCTCT 360
 Db 409 GCCAAACAATTTGGTGCACAGTTCATCGAAGAGTTCGAGTTTCAAAAGACAGCTCT 468
 Qy 361 ACCTCACTAGATGCATCATAGATTTCTTAATATATATATATATATATATATATATATATAT 420
 Db 469 ACCTCACTAGATGCATCATAGATTTCTTAATATATATATATATATATATATATATATATAT 528
 Qy 421 AATATTCAAGCTACTCTTCATGCTTTACATCTCTACTGATCTTCAAAAGTTTCAGAAATG 480

```

Db      529 AATATTCAGCTACCTCTCCATGTTTACATCCCTACTGATCTTCAAAAAGTTCAGAAATG 588
Qy      481 ATTGAGAAAGAGGATATGATCTGTTCTCTGTTTGTGAGACGCCATCAGTTTCGATGG 540
Db      589 ATTGAGAAAGAGGATATGATCTGTTTCTCTGTTTGTGAGACGCCATCAGTTTCGATGG 648
Qy      541 AGTGAATTCAGAAAGAGGATGTTGTTGAAGTACCGGACCTCTGAATTTAAATCCAGCTAAA 600
Db      649 AGTGAATTCAGAAAGAGGATGTTGTTGAAGTACCGGACCTCTGAATTTAAATCCAGCTAAA 708
Qy      601 CGGCTCTGTCGACAGCTGGATGAGAAATATATGAAATGGCTCATTTTATTTTGGT 660
Db      709 CGGCTCTGTCGACAGCTGGATGAGAAATATATGAAATGGCTCATTTTATTTTGGT 768
Qy      661 AAAAGACATTTGATAGAGATGGTTACTTTCAGGGTGG-AAAATGGCATACTACGAATG 719
Db      769 AAAAGACATTTGATAGAGATGGTTACTTTCAGGGTGG-AAAATGGCATACTACGAATG 828
Qy      720 CGAGCTGGACATAGTGTGATATAGATGTTGATATGATGCTTATTCAGAGCAAG 779
Db      829 CGAGCT-GAACATAGTGTGATATAGATGTTGATATGATGCTTATTCAGAGCAAG 887
Qy      780 AGTATTACATATGGCTA-TTTGGCAAAGAGAGCTTAAGGAAATATAAACTTTTGGTTTG 839
Db      888 AGTATTACATATGGCTA-TTTGGCAAAGAGAGCTTAAGGAAATATAAACTTTTGGTTTG 945
Qy      840 CAATATTGATGATGTCTCACCAGTGGCCACCATTTATGATACGGA 885
Db      946 CAATATTGATGATGTCTC-MCATGSSCACATTTATGATACGGA 990

RESULT 9
BM457667
LOCUS
DEFINITION
AGENCOURT 6412100 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5498829
5', mRNA sequence.
ACCESSION
BM457667.1 GI:18506707
VERSION
BM457667
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 904);
NIH-MGC Institutes of Health, Mammalian Gene Collection (MGC)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L14M12131 row: c column: 22
High quality sequence stop: 660.
Location/Qualifiers
1..904
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5498829"
/tissue_type="lymphoma, cell line"
/lab_host="DH1CB (phage-resistant)"
/clone_lib="NIH_MGC_85"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dr primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

FEATURES
source

```

ORIGIN

```

Query Match 61.4%; Score 801.4; DB 12; Length 904;
Best Local Similarity 96.8%; Pred. No. 1.2e-199;
Matches 855; Conservative 4; Mismatches 20; Indels 4; Gaps 4;
Qy      382 GAATTTCTTAATATATATGAGGTGACATTTGATAGAAATATTTCAAGCTACTTCTTCA 441
Db      13 GAATTTCTTAATATATATGAGGTGACATTTGATAGAAATATTTCAAGCTACTTCTTCA 72
Qy      442 TGTTCATCTCTACTGATCTTCAAAAAGTTGCAGAAATGATTCGAGAAAGATATGAT 501
Db      73 TGTTCATCTCTACTGATCTTCAAAAAGTTGCAGAAATGATTCGAGAAAGATATGAT 132
Qy      502 TCTGKTTTCTCTGTTGTGAGACGCCATCAGTTTCGATGGGTGAAATTCAGAAAGGAGTT 561
Db      133 TCTGTTTCTCTGTTGTGAGACGCCATCAGTTTCGATGGGTGAAATTCAGAAAGGAGTT 192
Qy      562 CQTGAAGTGCACCGAACCCTCTGAATTTAAATCCAGCTAAACGGCTCTGTCGACAAAGCTGG 621
Db      193 CQTGAAGTGCACCGAACCCTCTGAATTTAAATCCAGCTAAACGGCTCTGTCGACAAAGCTGG 252
Qy      622 GATGGAGAAATTAATCAAAAATGGCTCATTTTATTTTCTTAAAGACATTTGATAGATG 681
Db      253 GATGGAGAAATTAATCAAAAATGGCTCATTTTATTTTCTTAAAGACATTTTATAGATG 312
Qy      682 GGTATCTCGCAGGTGG-AAAATGGCATACTACGAAATCGAGCTGGAACATAGTGTGA 740
Db      313 GGTATCTCGCAGGTGG-AAAATGGCATACTACGAAATCGAGCTGGAACATAGTGTGA 371
Qy      741 TATAGATGTGGATATTTGATTTGGCTTATTCAGAGCAAGAGTATTAAGATATGGCTATTT 800
Db      372 TATAGATGTGGATATTTGATTTGGCTTATTCAGAGCAAGAGTATTAAGATATGGCTATTT 431
Qy      801 TGCAAAAGAGAGCTTAAGAAATAAACTTTTGGTTTGCATTAATTCGATGATGTCTCAC 860
Db      432 TGCAAAAGAGAGCTTAAGAAATAAACTTTTGGTTTGCATTAATTCGATGATGTCTCAC 491
Qy      861 CAATGCCACATTTATGATTCAGAGCAAGAAAGAAATAATATCTTATGATGTAAAGA 920
Db      492 CAATGCCACATTTATGATTCAGAGCAAGAAAGAAATAATATCTTATGATGTAAAGA 551
Qy      921 TGTATTGGGATAAGTTTATTAAGAAAGTGTATTGAGGTGAGGCTAATCTCAGAAAG 980
Db      552 TGTATTGGGATAAGTTTATTAAGAAAGTGTATTGAGGTGAGGCTAATCTCAGAAAG 611
Qy      981 GGCCTGTTTCAAGCAGACGCTGCTCTTTTAAACCTGGATTGCAAAATCGAGTCACTGT 1040
Db      612 GGCCTGTTTCAAGCAGACGCTGCTCTTTTAAACCTGGATTGCAAAATCGAGTCACTGT 671
Qy      1041 ATCAGACAAAGCTAGCAGTTGTAGATGAATGGAGAAAGAAATGGGCTGTGCTGGAAGA 1100
Db      672 ATCAGACAAAGCTAGCAGTTGTAGATGAATGGAGAAAGAAATGGGCTGTGCTGGAAGA 731
Qy      1101 AGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGTCTG-AGAGAGTGGGCTAA 1159
Db      732 AGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGTCTTGAAGAGAGTGGGCTAA 791
Qy      1160 GTGGGCTCTGCTGATGCTGTCTTCTAGCCAGAGAGGCTGTTCGATACATTTGCAAT 1219
Db      792 GTGGGCTCTGCTGATGCTGTCTTCTAGCCAGAGAGGCTGTTCGATACATTTGCAAT 851
Qy      1220 GTAAATGGT-GGGCGTGGTGCATCCGAGAAATTTGCAGAGCACCA 1261
Db      852 GTAAATGGTGGGCGCGTGGGCCATCCCGAGAAATTTGCCAGAGCA 894

```

RESULT 13

```

BQ226032
LOCUS
DEFINITION
AGENCOURT_7543829 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:606899
5', mRNA sequence.
BQ226032
ACCESSION

```

```

BQ226032
LOCUS
DEFINITION
AGENCOURT_7543829 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:606899
5', mRNA sequence.
BQ226032
ACCESSION

```

QY	481	ATTTCGAGAGAGAGAGATATGATTCCTGTTTCTCTGTTGTTGAGAGCCCATCAGTTTCGATGG	540
Db	490	ATTTCGAGAGAGAGAGATGATTCCTGTTTCTCTGTTGTTGAGAGCCCATCAGTTTCGATGG	540
QY	541	AGTGAATTCAGAAAGAGAGTTCCTGAGTGAACCGAACCTCTGAAATTAATAATCCAGCTAAA	600
Db	550	AGTGAATTCAGAAAGAGAGTTCCTGAGTGAACCGAACCTCTGAAATTAATAATCCAGCTAAA	609
QY	601	CGGCCTCGTCGACAGACTGGGATGAGGAATATATGAAATGCTCATTTTATTTTGTCT	660
Db	610	CGGCCTCGTCGACAGACTGGGATGAGGAATATATGAAATGCTCATTTTATTTTGTCT	669
QY	661	AAAAGACATTTGATAGAGATGGCTTACTTTCAGGGTGG-AAAATGGCATACTACGAAATG	719
Db	670	AAAAGACATTTGATAGAGATGGCTTACTTTCAGGGTGG-AAAATGGCATACTACGAAATG	729
QY	720	CGAGCTGGAAACATAGTGTGGATATAGATGT-GGATATTGATTGGCTTATTCAGAGCAAA	778
Db	730	CGAGCTGGAAACATAGTGTGGATATAGATGT-GGATATTGATTGGCTTATTCAGAGCAAA	788
QY	779	GAGTATTAAAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAATAATAAACTTTTGGTTT	838
Db	790	GAGTATTAAAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAATAATAAACTTTTGGTTT	849
QY	839	GCAAT 843	
Db	850	GCAAT 854	
RESULT 11			
AL538941		1201 bp mRNA linear EST 31-MAY-2003	
LOCUS		AL538941 Homo sapiens FETAL BRAIN Homo sapiens CDNA clone	
DEFINITION		CS0DF030YB08 5-PRIME, mRNA sequence.	
ACCESSION		AL538941	
VERSION		AL538941.2 GI:31263511	
KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		1 (bases 1 to 1201)	
AUTHORS		Li, W.B., Gruber, C., Jessee, J. and Polayes, D.	
TITLE		Full-length cDNA libraries and normalization	
JOURNAL		Unpublished (2001)	
COMMENT		On Feb 15, 2001 this sequence version replaced gi:12867704.	
FEATURES		source	
		1. .1201	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="CS0DF030YB08"	
		/tissue_type="FETAL BRAIN"	
		/dev_stage="fetal"	
		/clone_lib="Homo sapiens FETAL BRAIN"	
		/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."	
ORIGIN			
Query Match		61.2%; Score 798.6; DB 13; Length 881;	
Best Local Similarity		97.8%; Pred. No. 6.3e-199;	
Matches		826; Conservative 4; Mismatches 13; Indels 2; Gaps 2;	
QY	1	ATGACTCGGTGAGAGAGGGGGCGGCACCTCTCTCCAAACCGCGGGGGGAGCGCTCC	60
Db	10	ATGACTCGGTGAGAGAGGGGGCGGCACCTCTCTCCAAACCGCGGGGGGAGCGCTCC	69
QY	61	CGGGGGCGGGCGGCGAGCTCGAGCGCAACTCTCGCGGGCGGAGCGGCGAGGTGTGGAG	120
Db	70	CGGGGGCGGGCGGCGAGCTCGAGCGCAACTCTCGCGGGCGGAGCGGCGAGGTGTGGAG	129
QY	121	ARGCCCCCGCACCTGGCAGCCCTTAATTTCTGGCCCGGGGAGCGAGCAAGCAATCCCCCTG	180
Db	130	ARGCCCCCGCACCTGGCAGCCCTTAATTTCTGGCCCGGGGAGCGAGCAAGCAATCCCCCTG	189
QY	181	ARGAACATTAAAGACCTGGCGGGGGTCCCGCTCATTTGGCTGGGTCCTGGCGGGCGCTG	240
Db	190	ARGAACATTAAAGACCTGGCGGGGGTCCCGCTCATTTGGCTGGGTCCTGGCGGGCGCTG	249
QY	241	GATTTCAGGGGCGCTTCCAGAGTGTA-TGGGTTTCGACAGACCATGATGAATTTGAGAAATGTG	300
Db	250	GATTTCAGGGGCGCTTCCAGAGTGTA-TGGGTTTCGACAGACCATGATGAATTTGAGAAATGTG	309
QY	301	GCCAAACAATTTGGTGACAGAGTTCATCGAAGAGTTCTGAAGTTTCAAAGACAGCTCT	360
Db	310	GCCAAACAATTTGGTGACAGAGTTCATCGAAGAGTTCTGAAGTTTCAAAGACAGCTCT	369
QY	361	ACCTCACTAGATGCATCATAGAAATTTCTTAATTATTAATCAATGAGGKTGCATTTGATGGA	420
Db	370	ACCTCACTAGATGCATCATAGAAATTTCTTAATTATTAATCAATGAGGKTGCATTTGATGGA	429
QY	421	AATATTCAAGCTACTCTTCATGTTTACATCTCTACTGATCTTCAAAAGTTGCAGAAATG	480
Db	430	AATATTCAAGCTACTCTTCATGTTTACATCTCTACTGATCTTCAAAAGTTGCAGAAATG	489

QY 801 TGGCAAGAGAGAGCTTAAGGAAATAAAACCTTTTGGTTTGCATATTTGATGATGCTCTC 860
 Db 541 NNN 600
 QY 861 CAATGGCCCATTTATGATCAGGAGACCAAAAGAACTAATCTTATGATGTAAAGA 920
 Db 601 NNN 660
 QY 921 TGTATTGGGATAAGTTTATTAAGAAAGTGGTATTAGTGGTGGCTTAATCTCAGAAAG 980
 Db 661 NNN 720
 QY 981 GGCTCTTAAAGCAGCGCTGCTCTTTTAAACTGCGATTGCAAAATGGAAGTCAAGTGT 1040
 Db 721 GGCTCTTCAAGCAGCGCTGCTCTTTTAAACTGCGATTGCAAAATGGAAGTCAAGTGT 780
 QY 1041 ATCAGACAAAGCTAGCTGATGATGAATCGAGAAAGAAATGGGCTGTGCTGGAAGA 1100
 Db 781 ATCAGACAAAGCTAGCTGATGATGAATCGAGAAAGAAATGGGCTGTGCTGGAAGA 840
 QY 1101 AGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGGTGGAGAGAGTGGGCTTAAG 1160
 Db 841 AGTGGCATATCTTGGAAATGAAGTGTCTGATGAAGAGTGGTGGAGAGAGTGGGCTTAAG 900
 QY 1161 TGGCGCTCTGCTGATGCTGCTCTTACGCGCCGAGAGGCTGTGGATACATTTGCAATG 1220
 Db 901 TGGCGCTCTGCTGATGCTGCTCTTACGCGCCGAGAGGCTGTGGATACATTTGCAATG 960
 QY 1221 TAAATGCTGGCGCTGGTGGCCATCCGAGAAATTCGAGACACATTTGCCCTACTTAATGGA 1280
 Db 961 TAAATGCTGGCGCTGGTGGCCATCCGAGAAATTCGAGACACATTTGCCCTACTTAATGGA 1020
 QY 1281 AGTTAATAATTCATGCAAAAATAG 1305
 Db 1021 GGTTAATAATTCATGCAAAAATAG 1045

RESULT 13
 BG674331
 LOCUS 602620310F1 NCI_CGAP_Skn3 982 bp mRNA linear EST 01-MAY-2001
 DEFINITION mRNA sequence.
 ACCESSION BG674331
 VERSION BG674331.1 GI:13905727
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 982)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA
 Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM10593 row: c column: 04
 High quality sequence stop: 827.
 Location/Qualifiers
 1. 982
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4745787"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NCI_CGAP_Skn3"

/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 ORIGIN
 Query Match 57.6%; Score 751.6; DB 12; Length 982;
 Best Local Similarity 97.7%; Pred. No. 1.6e-186;
 Matches 824; Conservative 1; Mismatches 11; Indels 7; Gaps 6;
 466 AAAGTTTCAGAAATGATTTCGAGAGAGGATATGATCTCTGTTCTCTGTTCTGAGAGCGC 525
 Db 2 AAAGTGCAGAAATGATTTCGAGAGAGGATATGATCTCTGTTCTCTGTTCTGAGAGCGC 61
 QY 526 CATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAGTGACCGAACCTCTGAAT 585
 Db 62 CATCAGTTTCGATGGAGTGAATTCAGAAAGGAGTTCGTGAAGTGACCGAACCTCTGAAT 121
 QY 586 TTAATCCAGCTAAACGGGCTCTGCGACAGACTGGGATGGAGATTAATATGAAAATGCGC 645
 Db 122 TTAATCCAGCTAAACGGGCTCTGCGACAGACTGGGATGGAGATTAATATGAAAATGCGC 181
 QY 646 TCATTTTATTTTGTCTAAAGACATTTTCATAGAGATGGTCTTCTGAGGGTGG-AAAATG 704
 Db 182 TCATTTTATTTTGTCTAAAGACATTTTCATAGAGATGGTCTTCTGAGGGTGGAAAATG 241
 QY 705 GCATCTACGAAATGCGAGCTGGAAACATAGTGTGGATATAGATGTGGATATTTGATTTGCC 764
 Db 242 GCATCTACGAAATGCGAGCT-GAACATAGTGTGGATATAGATGTGGATATTTGATTTGCC 300
 QY 765 TATTCGACAGCAAGAGATTAATAGATATGCTATTTGGCAAGAGAGAGCTTAAGGAAT 824
 Db 301 TATTCGACAGCAAGAGATTAATAGATATGCTATTTGGCAAGAGAGAGCTTAAGGAAT 360
 QY 825 AAAAATTTTGGTTTGGCAATTTGATGATGCTCTCACCATGGCCACATTTATGATCAGG 884
 Db 361 AAAAATTTTGGTTTGGCAATTTGATGATGCTCTCACCATGGCCACATTTATGATCAGG 420
 QY 885 AGACCAAAAGAAATTAATCTTATGATGTAAAGATGCTATTGGGATAAGTTTATTTAAA 944
 Db 421 AGACCAAAAGAAATTAATCTTATGATGTAAAGATGCTATTGGGATAAGTTTATTTAAA 480
 QY 945 GAAAGTGGTATTGAGTGGAGGCTTAATCTCAGAAAGGCGCTGTTCBAAGCAGAGCGCTGC 1004
 Db 481 GAAAGTGGTATTGAGTGGAGGCTTAATCTCAGAAAGGCGCTGTTCBAAGCAGAGCGCTGC 540
 QY 1005 TTCTTTAAACCTGGATTGCAAAATGGAAGTCAAGTGTATCAGCAAGCTAGCAGTTGTAGA 1064
 Db 541 TTCTTTAAACCTGGATTGCAAAATGGAAGTCAAGTGTATCAGCAAGCTAGCAGTTGTAGA 600
 QY 1065 TGAATCGAGAAAGAAATGGGCTGTGCTGGAAAGAGTGGCATATCTTGGAAATGAAGT 1124
 Db 601 TGAATCGAGAAAGAAATGGGCTGTGCTGGAAAGAGTGG-ATATCTTGGAAATGAAGT 659
 QY 1125 GTCTGATGAGAGTGTCTGAGAGAGTGGGCTTAAGTGGGCTCTGCTGATGCTGCTGTC 1184
 Db 660 GTCTGATGAGAGTGTCTGAGAGAGTGGGCTTAAGTGGGCTCTGCTGATGCTGCTGTC 729
 QY 1185 --CTAGCCCAAGAGGCTGTGGATACATTTGCAAAATGTAATGGTGGCGCTGGTGCATC 1242
 Db 720 TACTGGCCCAAGAGGCTGTGGATACATTTGCAAAATGTAATGGTGGCGTGGTG-CATC 778
 QY 1243 CGAGAAATTTGCGAGACACATTTGCGCTCTACTAATGGAAGAAAGTAAATTAATCATGCCAAAA 1302
 Db 779 CGAGAAATTTGCGAGACACATTTG-CCTACTAATGGAAGAGCTTAATTAATTAATTCATGCCAAAA 837
 QY 1303 TAG 1305
 Db 838 TAG 840
 RESULT 14
 BX343798

```

LOCUS      BX343798      1201 bp      mRNA      linear      EST 02-MAY-2003
DEFINITION BX343798 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0D1023VA10 5-PRIME, mRNA sequence.
ACCESSION  BX343798
VERSION    BX343798.1 GI:30344204
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1201)
            Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
            Full-length cDNA libraries and normalization.
            Unpublished (2001)
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6598.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS1A1006ZE06QPI&cluster=6598.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Paraday Avenue Genoscope sequence ID : CS1A1006ZE06QPI.
FEATURES   Location/Qualifiers
            1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0D1023VA10"
                /tissue_type="PLACENTA COT 25-NORMALIZED"
                /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                /notes="1st strand cDNA was primed with a NotI-oligo (dr)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and cloned into the Not I and EcoR V
                sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query March 57.1%; Score 745.2; DB 13; Length 1201;
Best Local Similarity 97.7%; Pred. No. 8.5e-185;
Matches 823; Conservative 6; Mismatches 4; Indels 9; Gaps 7;
QY 1 ATGACTCGTGGAGAGGGGGCGCCACCTCCGCTCCACCGCGGGGGCGGACCTCC 60
DB 146 ATGACTCGTGGAGAGGGGGCGCCACCTCCGCTCCACCGCGGGGGCGGACCTCC 205
QY 61 CGGGCGCGCGCGCGGAGCTGCAGCGCAACTCTCGCGCGCGCGCGGCGGAGGTGTGGAG 120
DB 206 CGGGCGCGCGCGCGGAGCTGCAGCGCAACTCTCGCGCGCGCGGCGGAGGTGTGGAG 265
QY 121 AAGCCCCCGACCTGGCAGGCTTAATTCGCGCGGGAGGACAGCAAGGCAATCCCGCTG 180
DB 266 AAGCCCCCGACCTGGCAGGCTTAATTCGCGCGGGAGGACAGCAAGGCAATCCCGCTG 325
QY 181 AAGAACATTAGCACTGGCGGGGGTCCGCTCATTCGCTGGGTCCTCGCTGGCGGCTG 240
DB 326 AAGAACATTAGCACTGGCGGGGGTCCGCTCATTCGCTGGGTCCTCGCTGGCGGCTG 385
QY 241 GATTCAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTTGAGAAATGTG 300
DB 386 GATTCAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATTTGAGAAATGTG 445
QY 301 GCCAAACAATTTGGTGACAAAGTTTCATCGAAGAGTCTGTGAAGTTTCAAAAGACAGCTCT 360
DB 446 GCCAAACAATTTGGTGACAAAGTTTCATCGAAGAGTCTGTGAAGTTTCAAAAGACAGCTCT 505
QY 361 ACCTCATAGTGCATCATAGATTTCTTAATTAATATATATATAGGCTGATTTAGGA 420
DB 506 ACCTCATAGTGCATCATAGATTTCTTAATTAATATATATATAGGCTGATTTAGGA 565
QY 421 AATATTCAGCTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCAGAAATG 480

```

```

Db 566 AATATTCAGCTACTTCTTCATGTTTACATCTCTACTGATCTTCAAAAAGTTGCAGAAATG 625
QY 481 ATTGAGAGAGAGATGATGATTCGCTTTCTCTGTTGTGAGACCCCATCAGTTTCGATGG 540
DB 626 ATTGAGAGAGAGATGATGATTCGCTTTCTCTGTTGTGAGACCCCATCAGTTTCGATGG 685
QY 541 AATGAAATTCAGAAAGGAGTTCTGTAAGTACCCGAACTCTGAATTTAAATCCAGCTAAA 600
DB 686 AGTGAATTCAGAAAGGAGTTCTGTAAGTACCCGAACTCTGAATTTAAATCCAGCTAAA 745
QY 601 CGGCTCTGTCGACAAAGCTGGGATGGAATATATATGAAATGGCTCATTTATTTCCT 660
DB 746 CGGCTCTGTCGACAAAGCTGGGATGGAATATATATGAAATGGCTCATTTATTTCCT 805
QY 661 AATGACATTTGATGAGATGGGTACTTCAGAGGTGG-AAAATGGCATATCAGAAATG 729
DB 806 AATGACATTTGATGAGATGGGTACTTCAGAGGTGG-AAAATGGCATATCAGAAATG 865
QY 720 CGAGCTGGAACATAGTGTGGATATAGATGTGATATGATGCTTATGCGCTATTGCGAGCAAG 779
DB 866 CGAGCT-GAACATAGTGTGGATATAGATGTGATGATGCTTATGCGCTATTGCGAGCAAG 922
QY 780 AGTATTAAGATATGCTTATTTTGGCAAGGAGAACTTAAGGAAATAAACTTTTGGCTTG 839
DB 923 AGTATTAAGATATGCTTATTTTGGCAAGGAGAACTTAAGGAAATAAACTTTTGGCTTG 977
QY 840 CA 841
DB 978 CA 979

RESULT 15
BQ438805
LOCUS      BQ438805      874 bp      mRNA      linear      EST 24-MAY-2002
DEFINITION AGENCOURT 7809455 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6053483
            5', mRNA sequence.
ACCESSION  BQ438805
VERSION    BQ438805.1 GI:21177881
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 874)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            E-mail: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)
            RNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILML at:
            http://image.llnl.gov
            Plate: LLAM3310 row: j column: 12
            High quality sequence stop: 602.
FEATURES   Location/Qualifiers
            1..874
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6053483"
                /tissue_type="epithelioid carcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_70"
                /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.1 kb. Library constructed by Life
                Technologies."
ORIGIN

```


Query Match 56.5%; Score 737.8; DB 13; Length 874;
Best Local Similarity 97.5%; Pred. No. 6.6e-183;
Matches 787; Conservative 4; Mismatches 11; Indels 5; Gaps 4;

Qy 1 ATGGACTCGGTGGAGAGGGGGCCGCCACCTCCCTCTCCAAACCCGGGGGGGGGCGACCTCC 60
Db |||||||
Qy 69 ATGGACTCGGTGGAGAGGGGGCCGCCACCTCCCTCTCCAAACCCGGGGGGGGGCGACCTCC 128
Db |||||||
Qy 61 CGGGCCCGGGCCCGGAGCTCGAGGCCAACTCTCGGGCGGCGAGGGCGGAGGTGTGGAG 120
Db |||||||
Qy 129 CGGGCCCGGGCCCGGAGCTCGAGGCCAACTCTCGGGCGGCGAGGGCGGAGGTGTGGAG 188
Db |||||||
Qy 121 AAGCCCCCGACCTGGGCGAGCCCTAATCTGGGCCGGGAGGCGAGAAAGGCGATCCGCCCTG 180
Db |||||||
Qy 189 AAGCCCCCGACCTGGGCGAGCCCTAATCTGGGCCGGGAGGCGAGAAAGGCGATCCGCCCTG 248
Db |||||||
Qy 181 AAGAACATTAAAGCACTCGGGGGGTCCGCTCATTTGGCTGGGTCTCTGGTGGGGCCCTG 240
Db |||||||
Qy 249 AAGAACATTAAAGCACTCGGGGGGTCCGCTCATTTGGCTGGGTCTCTGGTGGGGCCCTG 308
Db |||||||
Qy 241 GATTCAGGGGCGCTTCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATGTG 300
Db |||||||
Qy 309 GATTCAGGGGCGCTTCAGAGTGTATGGTTTCGACAGACCATGATGAATTCAGAAATGTG 368
Db |||||||
Qy 301 GCCAAACAATTTGGTGGCAAGTTCAATCGAAGAGTTCGAAGTTTCAGAAAGTTCAGAAAGTCT 360
Db |||||||
Qy 369 GCCAAACAATTTGGTGGCAAGTTCAATCGAAGAGTTCGAAGTTTCAGAAAGTTCAGAAAGTCT 428
Db |||||||
Qy 361 ACCTCAGTAGTCCATCATAGAAATCTTAATATATATATATATATATATATATATATATAT 420
Db |||||||
Qy 429 ACCTCAGTAGTCCATCATAGAAATCTTAATATATATATATATATATATATATATATATAT 488
Db |||||||
Qy 421 AATATTCAGGTACT 480
Db |||||||
Qy 489 AATATTCAGGTACT 548
Db |||||||
Qy 481 ATTGAGAGAGAGATATATATCT 540
Db |||||||
Qy 549 ATTGAGAGAGAGATATATCT 608
Db |||||||
Qy 541 AGTGAATTCAGAAAGAGTTCGTGAGTGACCGAACCTCTGAATTTAAATCCAGCTAAA 600
Db |||||||
Qy 609 AGTGAATTCAGAAAGAGTTCGTGAGTGACCGAACCTCTGAATTTAAATCCAGCTAAA 668
Db |||||||
Qy 601 CGGCTCTCTCGACAGACTGGGATCGAGAAATATATGAAATGGCTCATTTTATTTTGT 660
Db |||||||
Qy 669 CGGCTCTCTCGACAGACTGGGATCGAGAAATATGAAATGGCTCATTTTATTTTGT 728
Db |||||||
Qy 661 AAAGACATTTGATAGAGATGGGTTACTTGCAGGTGG-AAATGGCATCTACGAAATG 719
Db |||||||
Qy 729 AAAGACATTTGATAGAGATGGGTTACTTGCAGGTGGAAATGGCTTACTTACGAAATG 788
Db |||||||
Qy 720 CGACTGGAACATAGTGTGGATATAGATGTGG-AAATGATTTGGCTTATTCAGAGCAA 777
Db |||||||
Qy 789 CGASCT-GAACAATAGTGGGATAGATGTGGATATGGATGGGCTTATTCAGAGCAA 847
Db |||||||
Qy 778 AGAGTATTAAAG-ATATGGCTATTTTGG 803
Db |||||||
Qy 848 AGAGTATTAAAGTATGGCTATTTTGG 874
Db |||||||

Search completed: May 11, 2004, 11:45:54
Job time : 3795.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw mode

Run on: May 11, 2004, 11:46:35 ; Search time 604.697 Seconds
(without alignments)
9776.195 Million cell updates/sec

Title: US-09-930-440B-3

Perfect score: 1305

Sequence: 1 atgagctcgttgagagagg.....ataattcatgccaaaataag 1305

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 8

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB3.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB3.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1303.4	99.9	1305	9	US-09-984-205-1
2	1303.4	99.9	1305	9	US-09-930-440B-3
3	1272.4	97.5	1376	13	US-10-342-887-1801
4	1272.4	97.5	1376	13	US-10-172-118-1801
5	1272.4	97.5	1770	15	US-10-037-270-360
6	1272.4	97.5	1770	16	US-10-117-722-360
7	1272.4	97.5	1800	9	US-09-925-301-534
8	686.8	52.6	738	16	US-10-012-697-1444
9	647.6	49.6	789	16	US-10-012-697-1369
10	627.4	48.1	745	16	US-10-012-697-394
11	620.2	47.5	748	16	US-10-012-697-569
12	617.4	47.3	708	16	US-10-012-697-109
13	562	43.1	686	16	US-10-012-697-1128
14	555.6	42.6	639	9	US-09-867-550-1193

15	400.6	30.7	408	10	US-09-918-995-36321	Sequence 36321, A
16	306.8	23.5	509	9	US-09-783-590-293	Sequence 293, App
17	217	16.6	263	9	US-09-960-352-9191	Sequence 9191, App
C 18	185.6	14.2	251	9	US-09-796-692-3842	Sequence 3842, App
C 19	185.6	14.2	251	15	US-10-040-862-3842	Sequence 3842, App
C 20	185.6	14.2	251	16	US-10-057-475B-3842	Sequence 3842, App
C 21	185.6	14.2	251	16	US-10-154-884B-3842	Sequence 3842, App
22	145.8	11.2	222	11	US-09-864-408A-3297	Sequence 3297, App
23	145.4	11.1	205	9	US-09-960-352-12271	Sequence 12271, A
C 24	138.2	10.6	478	9	US-09-864-761-2461	Sequence 2461, App
C 25	134	10.3	134	9	US-09-864-761-19192	Sequence 19192, A
26	96.4	7.4	190	10	US-09-822-846-374	Sequence 374, App
C 27	79.2	6.1	95	9	US-09-864-761-11953	Sequence 11953, A
C 28	77.4	5.9	397	9	US-09-864-761-15439	Sequence 15439, A
29	60	4.6	60	10	US-09-903-975-5141	Sequence 5141, App
30	50.6	3.9	1830121	15	US-10-329-960-1	Sequence 1, Appli
31	50.6	3.9	1830121	16	US-10-329-960-1	Sequence 1, Appli
32	50.2	3.8	495	13	US-10-335-977-2927	Sequence 2927, App
33	50.2	3.8	1296	15	US-10-156-761-3361	Sequence 3361, App
C 34	50.2	3.8	9025608	15	US-10-156-761-1	Sequence 1, Appli
35	45.8	3.5	7462	15	US-10-311-455-1757	Sequence 1757, App
36	45.8	3.5	580073	15	US-10-205-220-1	Sequence 1, Appli
37	44.6	3.4	40862	15	US-10-311-455-2046	Sequence 2046, App
38	42.4	3.2	2430	15	US-10-156-761-2531	Sequence 2531, App
39	42.4	3.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
40	42	3.2	627	13	US-10-282-122A-25195	Sequence 25195, A
41	41.6	3.2	119211	13	US-10-672-787-40	Sequence 40, Appl
42	41.6	3.2	3186778	13	US-10-027-632-174961	Sequence 174961, A
43	41.6	3.2	3186778	16	US-10-027-632-174961	Sequence 174961, A
44	41.4	3.2	5802	13	US-10-152-886-84	Sequence 84, Appl
45	41	3.1	13249	15	US-10-311-455-89	Sequence 89, Appl

ALIGNMENTS

RESULT 1

US-09-984-205-1

; Sequence 1, Application US/09984205

; Patent No. US20020317175A1

; GENERAL INFORMATION:

; APPLICANT: Coleman, Timothy A. et al.

; TITLE OF INVENTION: Human Glycosylation Enzymes

; FILE REFERENCE: PFS0501

; CURRENT APPLICATION NUMBER: US/09/984,205

; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: PCT/US00/05325

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 09/516,143

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/122,409

; PRIOR FILING DATE: 1999-03-02

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1305

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1302)

; NAME/KEY: SITE

; LOCATION: (397)..(399)

; OTHER INFORMATION: yac encodes Tyr or His

; NAME/KEY: SITE

; LOCATION: (406)..(408)

; OTHER INFORMATION: gkt encodes Gly or Val

; NAME/KEY: SITE

; LOCATION: (439)..(441)

; OTHER INFORMATION: yca encodes Ser or Pro

; NAME/KEY: SITE

; LOCATION: (505)..(507)

; OTHER INFORMATION: gkt encodes Gly or Val

US-09-984-205-1

Query	Match	99.9%	Score	1303.4	DB	9	Length	1305
Best Local Similarity	100.0%	Prod.	No.	0				
Matches	1305	Conservative	0	Mismatches	0	Indels	0	Gaps
QY	1	ATGACACTCGGTGGAGAAAGGGGGCGCCACCTCCGCTCTCCAAACCCGCGGGGGGACACCTCC	60					
DB	1	ATGACACTCGGTGGAGAAAGGGGGCGCCACCTCCGCTCTCCAAACCCGCGGGGGGACACCTCC	60					
QY	61	CGGGCGCGGGCGCGCGAAAGCTGCAGCGCAACTCTCCGCGCGGCGGAGGCGAGAGTGTGGAG	120					
DB	61	CGGGCGCGGGCGCGCGAAAGCTGCAGCGCAACTCTCCGCGCGGCGGAGGCGAGAGTGTGGAG	120					
QY	121	AAGCCCCCGCACCTTGGCAGACCTTAATCTGCCCCGGGAGGCGAGGCAATCCCCCTG	180					
DB	121	AAGCCCCCGCACCTTGGCAGACCTTAATCTGCCCCGGGAGGCGAGGCAATCCCCCTG	180					
QY	181	AAGAACATTAAGCACCTTGGCGGGGGTCCCGCTCATTTGGCTGGGTCTCTGGTGGCGCCCTG	240					
DB	181	AAGAACATTAAGCACCTTGGCGGGGGTCCCGCTCATTTGGCTGGGTCTCTGGTGGCGCCCTG	240					
QY	241	GATTCAGGGGCTTCCAGAGTGTATGGTTTCGACAGCAACCATGATGAAATTTGAGAAATGTG	300					
DB	241	GATTCAGGGGCTTCCAGAGTGTATGGTTTCGACAGCAACCATGATGAAATTTGAGAAATGTG	300					
QY	301	GCCAAACAAATTTGGTGCACAAAGTTCTATCGAAGAAAGTTCTGAAGTTTCTCAAAGACAGCTCT	360					
DB	301	GCCAAACAAATTTGGTGCACAAAGTTCTATCGAAGAAAGTTCTGAAGTTTCTCAAAGACAGCTCT	360					
QY	361	ACCTCACTAGATGCCATCTATGAGAAATTTCTTAATTAATGATGAGGTGACATGTAGGA	420					
DB	361	ACCTCACTAGATGCCATCTATGAGAAATTTCTTAATTAATGATGAGGTGACATGTAGGA	420					
QY	421	AATATTCAAGCTACTCTTCATGTTTACATCCCTACTGATCTTCAAAAGTTTCAGAAATG	480					
DB	421	AATATTCAAGCTACTCTTCATGTTTACATCCCTACTGATCTTCAAAAGTTTCAGAAATG	480					
QY	481	ATTTCGAGAAAGGATATGATTCGKTTTCTCTGTTGTGAGAGCCCATCAGTTTTCGATGG	540					
DB	481	ATTTCGAGAAAGGATATGATTCGKTTTCTCTGTTGTGAGAGCCCATCAGTTTTCGATGG	540					
QY	541	AGTGGAAATTCAGAAAGGAGTTCGTGAGTGCACCGAACCTCTGAAATTAATTCAGCTAAA	600					
DB	541	AGTGGAAATTCAGAAAGGAGTTCGTGAGTGCACCGAACCTCTGAAATTAATTCAGCTAAA	600					
QY	601	CGGCTCGTGCACAGACTGGGATGGAGAAATTAATGAAATGGCTCATTTATTTTGGCT	660					
DB	601	CGGCTCGTGCACAGACTGGGATGGAGAAATTAATGAAATGGCTCATTTATTTTGGCT	660					
QY	661	AAAGACATTTGATAGATGGTTCATCTCGAGGTGGAAATGGCATCTACGAAATGC	720					
DB	661	AAAGACATTTGATAGATGGTTCATCTCGAGGTGGAAATGGCATCTACGAAATGC	720					
QY	721	GAGCTGGAACATAGTGTGGATATAGATGTGATTTGATTTGGCTATTCAGAGCAAGA	780					
DB	721	GAGCTGGAACATAGTGTGGATATAGATGTGATTTGATTTGGCTATTCAGAGCAAGA	780					
QY	781	GTATTAAAGATATGGCTATTTTGGCAAGAGAGCTTTAAGAAATAAAACTTTTGGTTGC	840					
DB	781	GTATTAAAGATATGGCTATTTTGGCAAGAGAGCTTTAAGAAATAAAACTTTTGGTTGC	840					
QY	841	AATATTGATGGATGCTCTACCAATGCCCATTTTGTATCAGGACCCAAAGAAATA	900					
DB	841	AATATTGATGGATGCTCTACCAATGCCCATTTTGTATCAGGACCCAAAGAAATA	900					
QY	901	ATATCTTATGATGATAAAGATGCTATTTGGGATAAGTTTATTAAGAAAAAGTGTGTAG	960					
DB	901	ATATCTTATGATGATAAAGATGCTATTTGGGATAAGTTTATTAAGAAAAAGTGTGTAG	960					
QY	961	GTGAGGCTAATCTCAGAAAGGGCTTTTCAAGCAGACGCTGTCTTCTTTAAACATGAT	1020					
DB	961	GTGAGGCTAATCTCAGAAAGGGCTTTTCAAGCAGACGCTGTCTTCTTTAAACATGAT	1020					

RESULT 2

```

US-09-930-440B-3
; Sequence 3, Application US/09930440B
; Patent No. US20020142386A1
; GENERAL INFORMATION:
; APPLICANT: Betenbaugh et al.
; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
; FILE REFERENCE: PF509P2
; CURRENT APPLICATION NUMBER: US/09/930,440B
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/227,579
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: 09/516,793
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/169,624
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: 60/122,582
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; NAME/KEY: misc.feature
; LOCATION: (397)..(399)
; OTHER INFORMATION: The 'yat' at location 397..399 encodes amino acid
; NAME/KEY: misc.feature
; LOCATION: (406)..(438)
; OTHER INFORMATION: The 'gkt' at location 406..438 encodes amino acid
; NAME/KEY: misc.feature
; LOCATION: (439)..(441)
; OTHER INFORMATION: The 'yca' at location 439..441 encodes amino acid
; NAME/KEY: misc.feature
; LOCATION: (505)..(507)
; OTHER INFORMATION: The 'gkt' at location 505..507 encodes amino acid
US-09-930-440B-3

```

Query Match	99.9%	Score 1303.4;	DB 9;	Length 1305;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1305;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	ATGACCTCGGTGGAGAAAGGGGGCGGCCACTCGTCTCCAAACCCGGGGGGGACCGTCC	60
Db	1	ATGACCTCGGTGGAGAAAGGGGGCGGCCACTCGTCTCCAAACCCGGGGGGGACCGTCC	60
Qy	61	CGGGCGCGCGCGCGCGAAGCTCGAGCGCAACTCTCGGGGGCGCAGGGCGGAGGTGTGAG	120


```

Db 439 AATATCAAGTACTCTCCAGTGTACATCTTACATCTTCAAAAAGTGGAGAATG 498
Qy 481 ATTCGAGAGAGAGATATGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 540
Db 499 ATTCGAGAGAGAGATATGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 558
Qy 541 AGTGAATTCAGAAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 600
Db 559 AGTGAATTCAGAAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 618
Qy 601 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 660
Db 619 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 678
Qy 661 AAAAGACATTTGATAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 719
Db 679 AAAAGACATTTGATAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 738
Qy 720 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 779
Db 739 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 797
Qy 798 AGTATTAAGATATGCTATTTGGCAAGAGAGATCTGATCTGATCTGATCTGATCTG 839
Db 840 CAATATTCATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 899
Qy 858 CAATATTCATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 917
Qy 900 AATATCTTATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 959
Db 918 AATATCTTATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 977
Qy 960 GGTGAGGCTATCTCAGAAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1019
Db 978 GGTGAGGCTATCTCAGAAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1037
Qy 1020 TTGCAAAATGCAAGTCACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1079
Db 1038 TTGCAAAATGCAAGTCACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1097
Qy 1080 AATGGGCTGCTGCTGGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1139
Db 1098 AATGGGCTGCTGCTGGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1157
Qy 1140 CTTGAGAGAGATGAGGCTTAACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1199
Db 1158 CTTGAGAGAGATGAGGCTTAACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1217
Qy 1200 TGTGATGATCAATTTGCAAAAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1259
Db 1218 TGTGATGATCAATTTGCAAAAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1277
Qy 1260 CATTTGCTACTAATGAAAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1305
Db 1278 CATTTGCTACTAATGAAAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1323

```

RESULT 4

```

US-10-172-118-1801
; Sequence 1801, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vliet, Marc
; APPLICANT: Bernards, Rene

```

```

; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1801
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_018686
; DATABASE ENTRY DATE: 2001-06-18
; US-10-172-118-1801

```

```

Query Match 97.5%; Score 1272.4; DB 13; Length 1376;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

Qy 1 ATGGACTCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Db 19 ATGGACTCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 78
Qy 61 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 79 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 138
Qy 121 AAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 139 AAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 198
Qy 181 AAGAACATTAAGACACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 199 AAGAACATTAAGACACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258
Qy 241 GATTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 259 GATTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 318
Qy 301 GCCAAACAATTTGTTGCAACAAGATCTGAAAGATCTGAAAGATCTGAAAGATCTGAAAG 360
Db 319 GCCAAACAATTTGTTGCAACAAGATCTGAAAGATCTGAAAGATCTGAAAGATCTGAAAG 378
Qy 361 ACCTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
Db 379 ACCTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 438
Qy 421 AATATTCAGCTACTCTCTCATGTTTACATCTGATCTGATCTGATCTGATCTGATCTGATCT 480
Db 439 AATATTCAGCTACTCTCTCATGTTTACATCTGATCTGATCTGATCTGATCTGATCTGATCT 498
Qy 481 ATTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Db 499 ATTGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 558
Qy 541 AGTGAATTCAGAAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 600
Db 559 AGTGAATTCAGAAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 618
Qy 601 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 660
Db 619 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 678
Qy 661 AAAAGACATTTGATAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 719
Db 679 AAAAGACATTTGATAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 738
Qy 720 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 779
Db 739 CCGCTCTGCGAAGAGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 797
Qy 780 AGTATTAAGATATGCTATTTGGCAAGAGAGATCTGATCTGATCTGATCTGATCTGATCTG 839

```

```
Db 798 AGTAATAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAAATAAAACCTTTTGGTTTG 857
Qy 840 CAATATTGATGATGTCTCACCATATGGCCACATTTATGTATCAGGAGACCAAAAAGAAAT 899
Db 858 CAATATTGATGATGTCTCACCATATGGCCACATTTATGTATCAGGAGACCAAAAAGAAAT 917
Qy 900 AATATCTTATGATGTAAAGATGCTATTGGGATAGTTTNTTAAAGAAAGTGGTATTGA 959
Db 918 AATATCTTATGATGTAAAGATGCTATTGGGATAGTTTNTTAAAGAAAGTGGTATTGA 977
Qy 960 GGTGAGGCTTATCTCAGAAAGGCTGTTCAAAGCAGACGCTGTCTCTTTTAAACTGGA 1019
Db 978 GGTGAGGCTTATCTCAGAAAGGCTGTTCAAAGCAGACGCTGTCTCTTTTAAACTGGA 1037
Qy 1020 TTGCAAAATGGAAGTCAAGTGTATCAGACAAGCTAGCAGTTGTAGATGAATGGAGAAAAGA 1079
Db 1038 TTGCAAAATGGAAGTCAAGTGTATCAGACAAGCTAGCAGTTGTAGATGAATGGAGAAAAGA 1097
Qy 1080 AATGGGCTGTGCTGGAAAGAGTGGCATATCTTTGGAAATGAAGTGTCTGATGAGAGTG 1139
Db 1098 AATGGGCTGTGCTGGAAAGAGTGGCATATCTTTGGAAATGAAGTGTCTGATGAGAGTG 1157
Qy 1140 CTTGAAGAGAGTGGGCTTAAGTGGGCTGCTGCTGATGCTGTCTTCTACGCCCGAGAGGC 1199
Db 1158 CTTGAAGAGAGTGGGCTTAAGTGGGCTGCTGCTGATGCTGTCTTCTACGCCCGAGAGGC 1217
Qy 1200 TGTGATATCATTTGCAAAATGTAATGTGGGCTGGTGGCCATCCGAGAAATTTGCAGAGCA 1259
Db 1218 TGTGATATCATTTGCAAAATGTAATGTGGGCTGGTGGCCATCCGAGAAATTTGCAGAGCA 1277
Qy 1260 CATTGGCTCTAATGCAAAAGTTAATATTCATGSCCAAAATAG 1305
Db 1278 CATTGGCTCTAATGCAAAAGTTAATATTCATGSCCAAAATAG 1323
```

RESULT 5

```
US-10-037-270-360
Sequence 360, Application US/10037270
Publication No. US20030104529A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyang
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Tillinghast, John
APPLICANT: Emarac, Radoje T.
TITLE OF INVENTION: No. US20030104529A1e1 Nucleic Acids and
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/10/037,270
PRIOR FILING DATE: 2002-01-04
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 360
LENGTH: 1770
TYPE: DNA
ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: (105)..(1409)
US-10-037-270-360
```

Query Match 97.5%; Score 1272.4; DB 15; Length 1770;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

```
Qy 1 ATGGAAGTGGGAGAGGGGGGCGCCACCTCCGCTCCAAACCGGGGGGGGCGACCGTCC 60
Db 105 ATGGAAGTGGGAGAGGGGGGCGCCACCTCCGCTCCAAACCGGGGGGGGCGACCGTCC 164
Qy 61 CGGGGCGGGCGCGCGGAGCTGCAGCGCACTCTCGCGCGCGGCGAGGGCGGAGTGTGAG 120
Db 165 CGGGGCGGGCGCGCGGAGCTGCAGCGCACTCTCGCGCGCGGCGAGGGCGGAGTGTGAG 224
Qy 121 AAGCCCGCGACCTGCGGAGCCCTAATCTGCGCGCGGAGGAGCAGCAAGGCAATCCCTCG 180
Db 225 AAGCCCGCGACCTGCGGAGCCCTAATCTGCGCGCGGAGGAGCAGCAAGGCAATCCCTCG 284
Qy 181 AAGAACATTAAGCACTGCGGGGGTCCCGCTCATTTGCTGGTCTGGTCTGGTCTGGTCT 240
Db 285 AAGAACATTAAGCACTGCGGGGGTCCCGCTCATTTGCTGGTCTGGTCTGGTCTGGTCT 344
Qy 241 GATTGAGGGGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAAATTCAGAAATGTG 300
Db 345 GATTGAGGGGCTTCCAGAGTGTATGGTTTCGACAGACCATGATGAAATTCAGAAATGTG 404
Qy 301 GCCAAACAATTTGTGCAAGTTCATGCAAGAGTCTGAAAGTTTCAAAAAGACAGCTCT 360
Db 405 GCCAAACAATTTGTGCAAGTTCATGCAAGAGTCTGAAAGTTTCAAAAAGACAGCTCT 464
Qy 361 ACCTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 465 ACCTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 524
Qy 421 AATATCAAGCTACTTCTTCATGTTTACATCTACTACTACTACTACTACTACTACTACTACT 480
Db 525 AATATCAAGCTACTTCTTCATGTTTACATCTACTACTACTACTACTACTACTACTACTACT 584
Qy 481 ATTGGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Db 585 ATTGGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 644
Qy 541 AGTGAATTCAGAAAGAGTTCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 600
Db 645 AGTGAATTCAGAAAGAGTTCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 704
Qy 601 CGGCTCTGTCGACAAAGCTGGATGGAGATTTATGAAATGGCTCATTTTATTTTGTCT 660
Db 705 CGGCTCTGTCGACAAAGCTGGATGGAGATTTATGAAATGGCTCATTTTATTTTGTCT 764
Qy 661 AAGAGCATTTGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719
Db 765 AAGAGCATTTGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 824
Qy 720 CGAGCTGGAAACATAGTGTGGATATAGATGATGATGATGATGATGATGATGATGATGAT 779
Db 825 CGAGCT-GAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 883
Qy 780 AGTATTAAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAAATAAACTTTTGGTTTG 839
Db 884 AGTATTAAGATATGGCTATTTTGGCAAGAGAGCTTAAGGAAATAAACTTTTGGTTTG 943
Qy 840 CAATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 899
Db 944 CAATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1003
Qy 900 AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959
Db 1004 AATATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1063
```

960 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCCAGACGACGCTGCTCTCTTTAAACCTGGA 1019
1064 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCCAGACGACGCTGCTCTCTTTAAACCTGGA 1123
1020 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCAAGTGTAGATGAATGGAGAAAGA 1079
1124 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCAAGTGTAGATGAATGGAGAAAGA 1183
1080 AATGGGCTGTGCTGGAAGAGAGTGGCATATCTTGAATGAAGTGTCTGTGATGAAGAGTG 1139
1184 AATGGGCTGTGCTGGAAGAGAGTGGCATATCTTGAATGAAGTGTCTGTGATGAAGAGTG 1243
1140 CTTGAAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTGTCTCTAGCCGAGAGGC 1199
1244 CTTGAAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTGTCTCTAGCCGAGAGGC 1303
1200 TGTGGATCATTTGCAATGAATGAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1259
1304 TGTGGATCATTTGCAATGAATGAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1363
1260 CATTGGCTACTAATGGAAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1305
1364 CATTGGCTACTAATGGAAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1409

RESULT 6

US-10-117-722-363
; Sequence 360, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1:el Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,322
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,327
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pc_f1_genes Version 1.0
; SEQ ID NO 360
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (105)..(1409)
US-10-117-722-363

Query Match 97.5%; Score 1272.4; DB 16; Length 1770;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;
Cyt 1 ATGACTCGGTGGAGAGGGGGCGCCACCTCCGCTCTCCAAACCGCGGGGGCGACCGTCC 60
Ddb 105 ATGACTCGGTGGAGAGGGGGCGCCACCTCCGCTCTCCAAACCGCGGGGGCGACCGTCC 164
Cyt 61 CGGGGGCGGGCGGAGAGCTGACAGGCACTCTCGCGCGGCCCGAGGCGGAGGTGTGAG 120
Ddb 165 CGGGGGCGGGCGGAGAGCTGACAGGCACTCTCGCGCGGCCCGAGGCGGAGGTGTGAG 224
Cyt 121 AAGCCCGCGACCTGGGAGCCCTAATTTCTGCGCGGGGGGAGGAGCAAGGCTCCCGCTG 180
Ddb 225 AAGCCCGCGACCTGGGAGCCCTAATTTCTGCGCGGGGGGAGGAGCAAGGCTCCCGCTG 284

QY 181 AAGAACATTAAGCACTGGCGGGGTCCTCCCTCATTCGCTGGGTCCTGGGTCCTGGGTCCTGG 243
Db 285 AAGAACATTAAGCACTGGCGGGGTCCTCCCTCATTCGCTGGGTCCTGGGTCCTGGGTCCTGG 344
QY 241 GATTACAGGGGCTTCCGAGAGTGTATGGGTTTCCGACAGCCATGATGAATTAAGAAATGTG 300
Db 345 GATTACAGGGGCTTCCGAGAGTGTATGGGTTTCCGACAGCCATGATGAATTAAGAAATGTG 404
QY 301 GCCAAACAAATTTGGTGCACAGTTTCATCGAAGAGTTCTCAAGTTTCAAAAGACAGCTCT 360
Db 405 GCCAAACAAATTTGGTGCACAGTTTCATCGAAGAGTTCTCAAGTTTCAAAAGACAGCTCT 464
QY 361 ACCTGCTAGTGCATCATAGAAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 420
Db 465 ACCTGCTAGTGCATCATAGAAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 524
QY 421 AATATTCAAGTACTTCTCTCATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
Db 525 AATATTCAAGTACTTCTCTCATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 584
QY 481 ATTGAGAAAGAGGATATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
Db 585 ATTGAGAAAGAGGATATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 644
QY 541 ACTGAAATTCAGAAAGAGTTCGTGAGTGCACGAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
Db 645 ACTGAAATTCAGAAAGAGTTCGTGAGTGCACGAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 704
QY 601 CGGCTCTCTGCGCAAGACTGGGATGGAGAAATTTATATGAATTAATTAATTAATTAATTAAT 660
Db 705 CGGCTCTCTGCGCAAGACTGGGATGGAGAAATTTATATGAATTAATTAATTAATTAATTAAT 764
QY 661 AAGAGCAATTTGATAGAGAGTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 719
Db 765 AAGAGCAATTTGATAGAGAGTGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 824
QY 720 CGAGCTGGAACATAGTGTGATATAGATGATGATGATGATGATGATGATGATGATGATGATG 779
Db 825 CGAGCT - GAACATAGTGTGATATAGATGATGATGATGATGATGATGATGATGATGATGAT 883
QY 780 AGTATTAAAGATATGCTTATTTGGCAAGAGAGAGCTTAAGGAAATTAAGAAATTAAGAAAT 839
Db 884 AGTATTAAAGATATGCTTATTTGGCAAGAGAGAGCTTAAGGAAATTAAGAAATTAAGAAAT 943
QY 840 CAATATTGATGAGTGTCTCAACCAATGGCCACATTTATGATGATGATGATGATGATGATGATG 899
Db 944 CAATATTGATGAGTGTCTCAACCAATGGCCACATTTATGATGATGATGATGATGATGATGATG 1003
QY 900 AATATCTTATGATGATAAAGATGCTATTGGGATAGTTTATTAAGAAATTAAGAAATTAAG 959
Db 1004 AATATCTTATGATGATAAAGATGCTATTGGGATAGTTTATTAAGAAATTAAGAAATTAAG 1063
QY 960 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCAAGCAGACGCTGCTCTCTCTCTCTCTCTCTCT 1019
Db 1064 GGTGAGGCTAATCTCAGAAAGGCGCTGTTCAAGCAGACGCTGCTCTCTCTCTCTCTCTCTCT 1123
QY 1020 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 1079
Db 1124 TTGCAAAATGGAAGTCAAGTGTATCAGACAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 1183
QY 1080 AATGGGCTGTGCTGGAAGAGAGTGGCATATCTTGGAAATGAAGTGTCTGTGATGAAGTG 1139
Db 1184 AATGGGCTGTGCTGGAAGAGAGTGGCATATCTTGGAAATGAAGTGTCTGTGATGAAGTG 1243
QY 1140 CTTGAAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTCTCTCTCTCTCTCTCTCTCTCTCT 1199
Db 1244 CTTGAAGAGTGGGCTTAAGTGGGCTCTCTGCTGATGCTCTCTCTCTCTCTCTCTCTCTCTCT 1303
QY 1200 TGTGGATCATTTGCAATGAATGAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1259
Db 1304 TGTGGATCATTTGCAATGAATGAATGGTGGCGGTGGTGGCCATCCGAGAAATTTGCAGAGCA 1363
QY 1260 CATTGGCTACTAATGGAAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1305

Db 1364 CATTGGCTACTAATGGAAGGTTAATAATCATGCCCCAAATATAG 1409

RESULT 7

US-09-925-301-534

; Sequence 534, Application US/09925301

; Patent No. US20020052308A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: P106

; CURRENT APPLICATION NUMBER: US/09/925,301

; CURRENT FILING DATE: 2001-03-10

; PRIOR APPLICATION NUMBER: PCT/US00/05982

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; PRIOR FILING DATE: 1999-03-12

; NUMBER OF SEQ ID NOS: 1694

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 534

; LENGTH: 1800

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-301-534

Query Match 97.5%; Score 1272.4; DB 9; Length 1800;

Best Local Similarity 99.2%; Pred. No. 0;

Matches 1295; Conservative 4; Mismatches 5; Indels 2; Gaps 2;

QY	1	ATGACCTCGGTGGAGAGGGGGCGGCGCCACCTCGGTCTCCAAACCGCGGGGGCGACCGTCC	60
DB	120	ATGACCTCGGTGGAGAGGGGGCGGCGCCACCTCGGTCTCCAAACCGCGGGGGCGACCGTCC	179
QY	61	CGGGGCGGGCGGCGGAGCTGCGCGGCACTCTCGCGGCGGCGGCGGCGGCGGCGGCGG	120
DB	180	CGGGGCGGGCGGCGGAGCTGCGAGCGCACTCTCGCGGCGGCGGCGGCGGCGGCGGCGG	239
QY	121	AGCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	180
DB	240	AGCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	299
QY	181	AGAAACATTAAGCACCTGGCGGGGCTCCGCTCAATGGCTGGGCTCGCTGGCGGGCGGCGG	240
DB	300	AGAAACATTAAGCACCTGGCGGGGCTCCGCTCAATGGCTGGGCTCGCTGGCGGGCGGCGG	359
QY	241	GATTCAGGGGCGCTTCAGAGTGTATGGGTTCCAGACACCATGATGAATTTGAGAAATGTG	300
DB	360	GATTCAGGGGCGCTTCAGAGTGTATGGGTTCCAGACACCATGATGAATTTGAGAAATGTG	419
QY	301	GCCAAACATTTGGTSCAACAGTTTCATCGAAGAGTTCTGAAAGTTTCAAAAGACAGCTCT	360
DB	420	GCCAAACATTTGGTSCAACAGTTTCATCGAAGAGTTCTGAAAGTTTCAAAAGACAGCTCT	479
QY	361	ACCTCACTAGATGCCATCATAGAAATTTCTTAATTAATGAGGKTGCACATTTGAGGA	420
DB	480	ACCTCACTAGATGCCATCATAGAAATTTCTTAATTAATGAGGKTGCACATTTGAGGA	539
QY	421	AATATTCAGGCTACTTCTTCATGTTTACATGCTGATCTTCAAAAAGTTTCAGAAATG	480
DB	540	AATATTCAGGCTACTTCTTCATGTTTACATGCTGATCTTCAAAAAGTTTCAGAAATG	599
QY	481	ATTGAGAGAGAGGATCATGATCTCTGTTTCTGTTGTCAGAGCGCCATCAGTTTCAGATG	540
DB	600	ATTGAGAGAGAGGATCATGATCTCTGTTTCTGTTGTCAGAGCGCCATCAGTTTCAGATG	659
QY	541	AGTGAATTCAGAAAGGAGTTCGTGAAGTGAACCGACCTCTGAAATTTAAATCCAGTAA	600
DB	660	AGTGAATTCAGAAAGGAGTTCGTGAAGTGAACCGACCTCTGAAATTTAAATCCAGTAA	719
QY	601	CGGCTCTGTCAGACAGCTGGAGATGAGAAATTTATGAAATTTGCTCAATTTTGTCT	660
DB	720	CGGCTCTGTCAGACAGCTGGAGATGAGAAATTTATGAAATTTGCTCAATTTTGTCT	779

RESULT 8

US-10-012-697-1444

; Sequence 1444, Application US/10012697

; Publication No. US20030215803A1

; GENERAL INFORMATION:

; APPLICANT: Escobedo, Jaime

; APPLICANT: Garcia, Pablo Dominguez

; APPLICANT: Kassam, Altaf

; APPLICANT: Lamson, George

; APPLICANT: Scott, Beth

; APPLICANT: Drmanac, Radoje

; APPLICANT: Crkvenjakov, Radomir

; APPLICANT: Dickson, Mark

; APPLICANT: Drmanac, Snezana

; APPLICANT: Labat, Ivan

; APPLICANT: Leshkowitz, Dena

; APPLICANT: Kita, David

; APPLICANT: Garcia, Veronica

; APPLICANT: Jones, Lee William

; APPLICANT: Stache-Crain, Birgit

; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS

; FILE OF INVENTION: ISOLATED FROM HUMAN PROSTATE

; FILE REFERENCE: 2300-16252

; CURRENT APPLICATION NUMBER: US/10/012,697

; PRIOR FILING DATE: 2003-05-21

; PRIOR APPLICATION NUMBER: 60/254,648

; PRIOR FILING DATE: 2000-12-07

QY	661	AAAAGACATTTGATAGAGATGGGTTACTTGCAGGCTGG-AAAATGGCATACTACGAATG	719
DB	780	AAAAGACATTTGATAGAGATGGGTTACTTGCAGGCTGGAAAANTGSCATACTACGAATG	839
QY	720	CGAGCTGGAAACATAGTGTGGATATAGCTGTGGATATTGATTTGGCCCTATTTCAGAGCAAG	779
DB	840	CGAGCT-GAACATAGTGTGGATATAGCTGTGGATATTGATTTGGCCCTATTTCAGAGCAAG	898
QY	780	AGTATTAGATATGCTATTTTGGCAAGAGAGCTTAAAGAAATAAAACCTTTTGGTTG	839
DB	899	AGTATTAGATATGCTATTTTGGCAAGAGAGCTTAAAGAAATAAAACCTTTTGGTTG	958
QY	840	CAATATTGATGATGTCTCACCAATGGCCACATTTATGATCAGGAGACCAAAAAGAAAT	899
DB	959	CAATATTGATGATGTCTCACCAATGGCCACATTTATGATCAGGAGACCAAAAAGAAAT	1018
QY	900	AATATCTTATGATGTAAAGAGTCTATTGGGATAGTTTATTAAGAAAGTGGTATTGA	959
DB	1019	AATATCTTATGATGTAAAGAGTCTATTGGGATAGTTTATTAAGAAAGTGGTATTGA	1078
QY	960	GGTGAGGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTGCTCTTTTAAACTGGA	1019
DB	1079	GGTGAGGCTAATCTCAGAAAGGGGCTGTTCAAAGCAGACGCTGCTCTTTTAAACTGGA	1138
QY	1020	TTGCAAAATGGAAGTCTAGTGTATCAGACAACTAGCAGCTTGTAGATGAATGGAGAAAGA	1079
DB	1139	TTGCAAAATGGAAGTCTAGTGTATCAGACAACTAGCAGCTTGTAGATGAATGGAGAAAGA	1198
QY	1080	AATGGGCGCTGTGCTGGAAAGAGTGGCATATCTTGGAAATGAAGTCTGTGTAAGAGTG	1139
DB	1199	AATGGGCGCTGTGCTGGAAAGAGTGGCATATCTTGGAAATGAAGTCTGTGTAAGAGTG	1258
QY	1140	CTTGAAGAGAGTGGGCGCTAAAGTGGCGCTCTGCTGATGCTCTTCTACGCCAGAGAGGC	1199
DB	1259	CTTGAAGAGAGTGGGCGCTAAAGTGGCGCTCTGCTGATGCTCTTCTACGCCAGAGAGGC	1318
QY	1200	TGTTGGATACATTTGCAAAATGTAATGCTGGCGCTGGTGGCCATCCGAGAAATTTGCAGAGCA	1259
DB	1319	TGTTGGATACATTTGCAAAATGTAATGCTGGCGCTGGTGGCCATCCGAGAAATTTGCAGAGCA	1378
QY	1260	CATTTCGCTACTAATGGAAGGTTAATAATTATTCATGCGCAAAATAG	1305
DB	1379	CATTTCGCTACTAATGGAAGGTTAATAATTATTCATGCGCAAAATAG	1424

;; PRIOR APPLICATION NUMBER: 60/275,668
;; PRIOR FILING DATE: 2001-03-13
;; NUMBER OF SEQ ID NOS: 1568
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1444
;; LENGTH: 738
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-012-697-1444

Query Match 52.6%; Score 686.8; DB 16; Length 738;
Best Local Similarity 97.3%; Pred. No. 1.2e-182;
Matches 716; Conservative 3; Mismatches 15; Indels 2; Gaps 2;
QY 405 GGTGACATGTAGAGAAATATTCACCTACTCTCTCATCTTACATCCCTACTGATCTTCA 464
DB 3 GGTGACATGTAGAGAAATATTCACCTACTCTCTCATCTTACATCCCTACTGATCTTCA 62
QY 465 AAAAGTTGCAGAAATGANTTCGAGAGAGAGATATGATTCCTGTTTCTGTTGTGAGAG 524
DB 63 AAAAGTTGCAGAAATGANTTCGAGAGAGAGATATGATTCCTGTTTCTGTTGTGAGAG 122
QY 525 CCATCAGTTTCATGAGAGTGAATTCAGAAAGAGTCTGTAAGTGAACCGAACCTCTGAA 584
DB 123 CCATCAGTTTCATGAGAGTGAATTCAGAAAGAGTCTGTAAGTGAACCGAACCTCTGAA 182
QY 585 TTAAATCCAGCTAAACGGCTCGTCGACAGACTGGGATGGAGATATATGAAATGG 644
DB 183 TTAAATCCAGCTAAACGGCTCGTCGACAGACTGGGATGGAGATATATGAAATGG 242
QY 645 CTCATTTTATTTGCTAAAGACATTTGATGAGATGGTGTACTTGCAGGGTGG-AAAT 703
DB 243 CTCATTTTATTTGCTAAAGACATTTGATGAGATGGTGTACTTGCAGGGTGGAAAT 302
QY 704 GGCATACCTAGAAATGCGAGCTGGACATAGTGTGATATGATGTGATATGATGTC 763
DB 303 GGCATACCTAGAAATGCGAGCT-GAACATAGTGTGATATGATGTGATATGATGTC 361
QY 764 CTATTGCAGAGCAAGAGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAA 823
DB 362 CTATTGCAGAGCAAGAGTATTAGATATGGCTATTTTGGCAAGAGAGAGCTTAAGGAA 421
QY 824 TAAACTTTTGGTTTGCATATATGATGATGTCTCACCATGGCCACATTTATGATCAG 883
DB 422 TAAACTTTTGGTTTGCATATATGATGATGTCTCACCATGGCCACATTTATGATCAG 481
QY 884 GAGACCAAAAGAAATATATCTTATGATGTAAAGATGCTATTGGGATATGTTTAA 943
DB 482 GAGACCAAAAGAAATATATCTTATGATGTAAAGATGCTATTGGGATATGTTTAA 541
QY 944 AGAAAGTGGTATTGAGGTAGGCTTAATCTCAGAAAGGGCTGTTCAGAGCAGAGCTGT 1003
DB 542 AGAAAGTGGTATTGAGGTAGGCTTAATCTCAGAAAGGGCTGTTCAGAGCAGAGCTGT 601
QY 1004 CTTCTTAAACTGGATGCAAAATGGAAGTCTGTATCAGCAAGCTACAGTTGTAG 1063
DB 602 CTTCTTAAACTGGATGCAAAATGGAAGTCTGTATCAGCAAGCTACAGTTGTAG 661
QY 1064 ATGAATGAGAAAGAAATGGCTGTCTGGAAGAGTGGCATATCTTGGAAATGAG 1123
DB 662 ATGAATGAGAAAGAAATGGCTGTCTGGAAGAGTGGCATATCTTGGAAATGAG 721
QY 1124 TGCTGTAGAGAGTG 1139
DB 722 TACTTGTAGGAAATG 737

RESULT 9
US-10-012-697-1369
; Sequence 1369, Application JS/10012697
; Publication No. US20030215803A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime

;; APPLICANT: Garcia, Pablo Dominguez
;; APPLICANT: Kassam, Altaf
;; APPLICANT: Lamson, George
;; APPLICANT: Scott, Beth
;; APPLICANT: Drmanac, Radoje
;; APPLICANT: Crkvenjakov, Radomir
;; APPLICANT: Dickson, Mark
;; APPLICANT: Drmanac, Snezana
;; APPLICANT: Labat, Ivan
;; APPLICANT: Leshkowitz, Dena
;; APPLICANT: Kita, David
;; APPLICANT: Garcia, Veronica
;; APPLICANT: Jones, Lee William
;; APPLICANT: Stache-Crain, Birgit
;; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
;; FILE REFERENCE: 2300-16252
;; CURRENT APPLICATION NUMBER: US/10/012,697
;; PRIOR FILING DATE: 2003-01-21
;; PRIOR APPLICATION NUMBER: 60/254,648
;; PRIOR FILING DATE: 2000-12-07
;; PRIOR APPLICATION NUMBER: 60/275,668
;; PRIOR FILING DATE: 2001-03-13
;; NUMBER OF SEQ ID NOS: 1568
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 1369
;; LENGTH: 789
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
;; NAME/KEY: misc feature
;; LOCATION: 2, 9, 11, 23, 33, 756, 771
;; OTHER INFORMATION: n = A,T,C or G
US-10-012-697-1369

Query Match 49.6%; Score 647.6; DB 16; Length 789;
Best Local Similarity 94.8%; Pred. No. 1.4e-171;
Matches 729; Conservative 5; Mismatches 22; Indels 13; Gaps 6;
QY 150 GGCCCGGGAGGAGCAGCAAGGSCATCCCTCGAAGAACATTAAGCACCTGGCGGGGTC 209
DB 14 GGCCCGGGAGGAGCAGCAAGGSCATCCCGGGAGAACATTAAGCACCTGGCGGGGTC 73
QY 210 GCTCATTTGGTGGTCTCTGCTGGCCCTGGATTCAGGGGCTTCCAGAGTGTATGGT 269
DB 74 GCTCATTTGGTGGTCTCTGCTGGCCCTGGATTCAGGGGCTTCCAGAGTGTATGGT 133
QY 270 TTGACAGACCATGATGAATTTGAGATGTGGCCAAACAATTTGGTGGACAAAGTTCT 329
DB 134 TTGACAGACCATGATGAATTTGAGATGTGGCCAAACAATTTGGTGGACAAAGTTCT 193
QY 330 AAGAAGTCTGAAGTTTCBAAGACAGCTCTACTCTACTAGATCCCATAGATTTCT 389
DB 194 AAGAAGTCTGAAGTTTCBAAGACAGCTCTACTCTACTAGATCCCATAGATTTCT 253
QY 390 TAATTATYATATGAGGKTGACATTTAGGAATATTTCAAGTACTTCTTCATCTTACA 449

Db	254	TATATTATCATTAAGGTTCACATTGTAGGAATAATCAAGCTACTTCTCCCATGTTTACA	313
Qy	450	TCCTACTGATCTTCAAAAAGTTGCAGAAATGATTCAGAGAAGAGGATATGATCTCGKTTT	509
Db	314	TCCTACTGATCTTCAAAAAGTTGCAGAAATGATTCAGAGAAGAGGATATGATCTCGT	373
Qy	510	CTCTGTTGTGTGAGCCCATCAGTTTCGATGTGAGTGAATTCAGAAAGGAGTTCGTGAAGT	569
Db	374	CTCTGTTGTGTGAGCCCATCAGTTTCGATGTGAGTGAATTCAGAAAGGAGTTCGTGAAGT	433
Qy	570	GACCGAACTCTGTAATTTAAATCCAGCTAAACGGCTCTGTCCACAAGACTCGGATGGAGA	629
Db	434	GACCGAACTCTGTAATTTAAATCCAGCTAAACGGCTCTGTCCACAAGACTCGGATGGAGA	493
Qy	630	ATTATATGAANAATGGCTCATTTTATTTTGCTTAAAGACATTTGATAGAGATGGGTACTTT	689
Db	494	ATTATATGAANAATGGCTCATTTTATTTTCTTAAAGACATTTGATAGAGATGGGTACTTT	553
Qy	690	GCAGGTGG-AAAATGGCATACTACCAATGCGAGCTGGACATAGTGTGGNATAGATG	748
Db	554	GCAGGTGGAAAATGGCATACTACCAATGCGAGCT-GAACATAGTGTGGNATAGATG	612
Qy	749	TGGATATTGATGGCTATTTCGAGAGCAAGAGTATTAAAGATATGGCTATTTTGGCAAG	808
Db	613	TGGATATTGATGGCTATTTCGAGAGCAAGAGTATTAAAGATATGGCTATTTTGGCAAG	672
Qy	809	AGAAGCTTAAGGAATAAACTTTTGGTTTGCATATTATGATGGATGCTCCACCATGGCC	868
Db	673	AGAAGCTTAAGGAA--TAAACTTTGGTT--GCATATTATGATGGATGCTCCACCATGGCCC	728
Qy	869	ACATTTATGATTCAGGAGACCAAAAAGAAATAATATCTTATGATATAA	917
Db	729	A----TTTGTATCAGGAGACCA---AGAAATATNTCTTATGATGTAA	773

RESULT 10

```

US-10-012-597-394
; Sequence 394, Application US/10012697
; Publication No. US2003021583A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Scott, Beth
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
; TITLE OF INVENTION: ISOLATED FROM HUMAN PROSTATE
; FILE REFERENCE: 2300-16252
; CURRENT APPLICATION NUMBER: US/10/012,697
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/254,648
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/275,668
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 394
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2, 9, 11, 23, 33, 627, 652, 673, 727, 731, 735

```



```
QY 310 TTTGGTGCACAGTTTCATCGAAGAGTTCTGAAGTTTCAAAGACAGCTCTACCTCACC'A 369
Db 124 TTTGGTGCACAGTTTCATCGAAGAGTTCTGAAGTTTCAAAGACAGCTCTACCTCACC'A 183
QY 370 GATGCCATCAVAGAAATTTCTTAATTAATVATCAATGAGGKTGACATTTGAGGAATATTCAA 429
Db 184 GATGCCATCAVAGAAATTTCTTAATTAATCAATGAGGTTGACATTTGAGGAATATTCAA 243
QY 430 GCTACTTCTYCATGTTTACATCTCTACTGATCTTTCAAAAAGTTTGCAGAAATGATCGAGRA 489
Db 244 GCTACTTCTCATGTTTACATCTCTACTGATCTTTCAAAAAGTTTGCAGAAATGATTCGAGAA 303
QY 490 GAAGATATGATTTCTGKTTTCTCTGTTGAGAGCCCATCAGTTTCGATGAGTGAATTT 549
Db 304 GAAGATATGATTTCTGTTTCTCTGTTGAGAGCCCATCAGTTTCGATGAGTGAATTT 363
QY 550 CAGAAAAGGAGTTCGTGAAGTGACCGAAGCTCTGAAATTTAAATCCAGCTAAACGGGCTCGT 609
Db 364 CAGAAAAGGAGTTCGTGAAGTGACCGAAGCTCTGAAATTTAAATCCAGCTAAACGGGCTCGT 423
QY 610 CGAAGAGCTGGATGGAGAAATATATGAAAATGCTCATTTATTTTGGTAAAGACAT 669
Db 424 CGAAGAGCTGGATGGAGAAATATATGAAAATGCTCATTTATTTTGGTAAAGACAT 483
QY 670 TTGATAGAGAT-GGGTACTTTCAGGGTGGAAAATGGCATACTACGAAATCGAGCTCGA 728
Db 484 TTGATAGAGATGGGTACTTTCAGGGTGGAAAATGGCATACTACGAAATCGAGCTCGA 542
QY 729 ACATAGTGTGATATGATGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 788
Db 543 ACATAGTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602
QY 789 ATATGGCTATTTTGGCAAGAGAGAG 813
Db 503 NIGGCTNTTTTGCRAANAGCTAAG 627
```

RESULT 14

```
US-09-867-550-1193
; Sequence 1193, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Foad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1193
; LENGTH: 639
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(154)
; OTHER INFORMATION: Wherein any n is one of a or t or c or g
US-09-867-550-1193
```

```
Query Match 42.6%; Score 555.6; DB 9; Length 639;
Best Local Similarity 94.8%; Pred. No. 1.1e-145;
Matches 580; Conservative 4; Mismatches 22; Indels 6; Gaps 1;

QY 1 ATGACATCGGTGGAGAGGGGGCCGACCTCGCTCTCCACCCCGGGGGCGACCGTCC 60
Db 34 ATGACATCGGTGGAGAGGGGGCCGACCTCGCTCTCCACCCCGGGGGCGACCGTCC 93
```

```
QY 61 CGGGCGCGGCGCGCGAGCTCGAGCGCAACTCTCCGCGCGCGCGCGCGAGTGTGGAG -20
Db 94 CGGGCGCGGCGCGCGAGCTCGAGCG-----ACTCTCGCGCGCGCGCGAGTGTGGAG -147
QY 121 AAGCCCCCGCACCTCGSAGCCCTTAATTTCTGGCCCCGGGAGGCGAGCAAAAGGCATCCCCCTG 180
Db 148 AAGCCCNCGACCTCGSAGCCCTTAATTTCTGGCCCCGGGAGGCGAGCAAAAGGCATCCCCCTG 207
QY 181 AAGAACATTAAAGCACTCGCGGGGTCCTCCGCTCATTTGGCTGGGTCTCTGGTGGGSCCTG 240
Db 208 AAGAACATTAAAGCACTCGCGGGGTCCTCCGCTCATTTGGCTGGGTCTCTGGTGGGSCCTG 267
QY 241 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATGAGAAATG 300
Db 268 GATTGAGGGGCTTCCAGAGTGTATGGGTTTCGACAGACCATGATGAATGAGAAATG 327
QY 301 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 360
Db 328 GCCAAACAATTTGGTGCAAGTTTCATCGAAGAGTTCTGAAGTTTCAAAAGACAGCTCT 387
QY 361 ACCTCAGTAGTGCATCATAGAAATTTCTTAATTAATATATAGAGTGCATTTAGGA 420
Db 388 ACCTCAGTAGTGCATCATAGAAATTTCTTAATTAATATATAGAGTGCATTTAGGA 447
QY 421 AATATTCAAGTACTTCTTCATGTTTACTCTCTACTGATCTTCAAAAAGTTCGAGAAATG 480
Db 448 AATATTCAAGTACTTCTTCATGTTTACTCTCTACTGATCTTCAAAAAGTTCGAGAAATG 507
QY 481 ATTCGAGAAGAGGATATGATTTCTGTTTCTCTGTTGTGAGAGCGCATCAGTTTCGATGG 540
Db 508 ATTCGAGAAGAGGATATGATTTCTGTTTCTCTGTTGTGAGAGCGCATCAGTTTCGATGG 567
QY 541 AGTGAAATTCAGAAAGAGGAGTTCTGTAAGTGACCGAACCTCTGAATTTAAATCCAGCTAAA 600
Db 568 AGTGAAATTCAGAAAGAGGAGTTCTGTAAGTGACCGAACCTCTGAATTTAAATCCAGCTAAA 627
QY 601 CGGCTCTGTCGA 612
Db 628 CGGCTCTGTCGA 639
```

RESULT 15

```
US-09-918-995-36321
; Sequence 36321, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36321
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36321
```

```
Query Match 30.7%; Score 400.6; DB 10; Length 408;
Best Local Similarity 99.0%; Pred. No. 3.9e-102;
Matches 403; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 871 ATTATCTATCAGAGACCAAAAGAAATATATCTTATCATGTATAAGATGCTATTGGG 930
Db 1 ATTATCTATCAGAGACCAAAAGAAATATATCTTATCATGTATAAGATGCTATTGGG 60
QY 931 ATAAGTTTATTAAGAAAGTGGTATTGAGGTGAGGTAAATCTCAGAAAAGGCGCTGTCA 990
```


Db 61 ATAGTTATTAAAGAAAGTGGTATTGAGTGAGGCTAATCTCGAAAGGGCCTTTCA 120
QY 991 AAGCAGACGCTGTCTTCTTTAAACTGGATTGCAAAATGGAAGTCAGTGTATCAGACAAG 1050
Db 121 AAGCAGACGCTGTCTTCTTTAAACTGGATTGCAAAATGGAAGTCAGTGTATCAGACAAG 180
QY 1051 CTAGCAGTTGTAGTGAATGGAGAAAGAAATGGGCTGTGCTGGAAGAGTGGCATAT 1110
Db 181 CTAGCAGTTGTAGTGAATGGAGAAAGAAATGGGCTGTGCTGGAAGAGTGGCATAT 240
QY 1111 CTTGGAATGAAAGTGTCTCATGAAGAGTCTTGAAGAGAGTGGGCTTAAGTGGCGCTCT 1170
Db 241 CTTGGAATGAAAGTGTCTCATGAAGAGTCTTGAAGAGAGTGGGCTTAAGTGGCGCTCT 300
QY 1171 GCTGATGCTGTTCCTACGCCAGAGGCTGTGGATACATTTGCAAAATGTAATGGTGGC 1230
Db 301 GCTGATGCTGTTCCTACGCCAGAGGCTGTGGATACATTTGCAAAATGTAATGGTGGC 360
QY 1231 CGTGGTGGCATCCGAGATTTGCAGAGCACATTTGCCTACTAATGGA 1277
Db 361 CGTGGTGGCATCCGAGATTTGCAGAGCACATTTGCCTACTAATGGA 407

Search completed: May 12, 2004, 12:37:19
Job time : 624.697 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:43 ; Search time 61.1848 Seconds
(without alignments)
2004.184 Million cell updates/sec

Title: US-09-930-440B-4

Perfect score: 2266

Sequence: 1 MDSVEGAATSVSNPRGRPS.....REFAHICLLMEKVNNSQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 3%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2258	99.6	434	3	Aay96100 Human CMP
2	2258	99.6	434	3	Aay90351 Human Gly
3	2258	99.6	434	4	Aab84682 Nucleotid
4	2258	99.6	434	6	Aao26545 Human CMP
5	2188	96.6	434	3	Aab28673 Human car
6	2188	96.6	434	4	Aam39319 Human pro
7	2188	96.6	434	4	Aab95213 Human pro
8	2188	96.6	448	3	Aab43931 Human can
9	2181	96.2	434	3	Aab43165 Human ORF
10	2149	94.8	461	4	Aam41105 Human pol
11	1173	51.8	245	7	ADB82754 Human pro
12	945	41.7	202	5	ABP64227 Human ORF
13	419	18.5	144	4	AAU29738 Novel hum
14	339	15.0	74	5	ABP32676 Human ORF
15	241	10.6	44	4	AAM16058 Peptide #
16	241	10.6	44	4	ABB35050 Peptide #
17	241	10.6	44	4	ABB29866 Peptide #
18	241	10.6	44	4	ABB20462 Protein #
19	241	10.6	44	4	Aam68236 Human bon
20	241	10.6	44	4	AAM55865 Human bra
21	241	10.6	44	4	ABG49887 Human liv
22	241	10.6	44	4	AAM03785 Peptide #
23	228	10.1	188	6	ABM69485 Phototab
24	225.5	10.0	169	3	Aab18422 Amino aci
25	217	9.6	228	3	Aay50798 N. mening

26	210.5	9.3	178	6	ABP80888	Abp80888 N. gonorr
27	208.5	9.2	410	3	AAy68969	AAy68969 Cps2r whi
28	198.5	8.8	413	5	ABP26813	ABP26813 Streptoco
29	177.5	7.8	209	6	ABU33455	ABU33455 Protein e
30	169.5	7.5	221	3	AAy97209	AAy97209 Campyloba
31	169.5	7.5	221	6	ABJ18499	ABJ18499 Campyloba
32	159	7.0	180	6	ADA35085	ADA35085 Acinetoba
33	154	6.8	1012	4	ABG18198	ABG18198 Novel hum
34	154	6.8	1012	4	ABG28888	ABG28888 Novel hum
35	145	6.4	536	6	ABJ18487	ABJ18487 Campyloba
36	145	6.4	536	6	ABJ18490	ABJ18490 Campyloba
37	138.5	6.1	498	4	ABG25738	ABG25738 Novel hum
38	137.5	6.1	165	2	AAW20503	AAW20503 H. pylori
39	128	5.6	467	2	AAZ21577	AAZ21577 HCV CKS-N
40	128	5.6	467	2	AAZ33644	AAZ33644 HCV CKS-N
41	128	5.6	467	2	AAZ33606	AAZ33606 HCV CKS-N
42	128	5.6	467	2	AAZ33586	AAZ33586 HCV CKS-N
43	128	5.6	467	4	AAZ69005	AAZ69005 HCV recom
44	128	5.6	541	2	AAZ21567	AAZ21567 HCV CKS-N
45	128	5.6	541	2	AAZ33634	AAZ33634 HCV CKS-N

ALIGNMENTS

RESULT 1
AAZ96100
ID AAY96100 standard; protein; 434 AA.
XX
AC AAY96100;
XX
DT 19-DEC-2000 (first entry)
XX
DE Human CMP-sialic acid synthetase.
XX
KW CMP-sialic acid synthetase; human; sialylation; glycoprotein;
KW plasminogen; transferrin; thyrotropin; Na+,K+-ATPase.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT Misc-difference 133 /note= "encoded by YAT"
FT Misc-difference 136 /note= "encoded by GKT"
FT Misc-difference 147 /note= "encoded bu YCA"
FT Misc-difference 169 /note= "encoded by GKT"
XX
WC200052135-A2.
XX
OB-SEP-2000.
XX
OI-MAR-2000; 2000WO-US005313.
XX
O2-MAR-1999; 99US-0122582P.
XX
O3-DEC-1999; 99US-0169624P.
XX
{HUMA-} HUMAN GENOME SCI INC.
{UYJO } UNIV JOHNS HOPKINS.
{UYWY-} UNIV WYOMING.
XX
Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
WPI; 2000-572178/53.
DR N-PSDB; AAR50568.
XX
Recombinant production of sialylated glycoproteins using cells in which
the expression of enzymes, e.g. sialic acid synthetase, involved in the
PT sialylation reaction has been altered.
XX
Claim 20; Page 102-103; 144pp; English.
PS

XX The present sequence is that of human CMP-sialic acid synthetase. The
 CC invention provides methods and recombinantly engineered cells for
 CC producing glycoproteins having sialylated oligosaccharides. The methods
 CC involve altering the expression of enzymes involved in carbohydrate
 CC processing. A claimed cell producing sialylated glycoprotein at above
 CC endogenous levels expresses at least 1 (preferably human) enzyme selected
 CC from GlcNAc-2 epimerase, an enzyme catalyzing the conversion of UDP-
 CC GlcNAc to ManNAc, sialic acid synthetase, aldolase, CMP-sialic acid
 CC synthetase and CMP-sialic acid transporter at above endogenous levels.
 CC Endogenous N-acetylglucosaminidase activity may be suppressed. A claimed
 CC method for manipulating glycoprotein in an insect cell comprises
 CC enhancing the expression of 1 of the above enzymes, and a claimed method
 CC for producing sialylated glycoproteins involves expressing a heterologous
 CC protein (especially sialylated glycoproteins) in an insect cell, plant and
 CC thrytrotropin) in the insect cell, yeast, insect, fungal, plant and
 CC bacterial host cells can be engineered to produce new forms of sialylated
 CC glycoproteins, higher concentrations of sialylated glycoproteins and/or
 CC elevated concentrations of donor substrates (e.g. nucleotide sugars)
 CC required for sialylation
 XX
 SQ Sequence 434 AA;

Query Match 99.6%; Score 2258; DB 3; Length 434;
 Best Local Similarity 100.0%; Pred. No. 2.2e-223;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGATSVSNRGRPSRGRPKLQNSRGGQGRGVERPPHAAALILARGGSGKGL 60
 Db 1 MDSVEKGATSVSNRGRPSRGRPKLQNSRGGQGRGVERPPHAAALILARGGSGKGL 60
 QY 61 KXKHLAIVPLIGVLRALDSGAFQSVWVSTDDHEIENVAQFQVHRRSSEVSKDSS 120
 Db 61 KXKHLAIVPLIGVLRALDSGAFQSVWVSTDDHEIENVAQFQVHRRSSEVSKDSS 120
 QY 121 TSLDAIIEFLNXXEXDIVGNIQATXCLHPDQLQKVAEMIREEGYDSXFSVVRHQRW 180
 Db 121 TSLDAIIEFLNXXEXDIVGNIQATXCLHPDQLQKVAEMIREEGYDSXFSVVRHQRW 180
 QY 181 SEIQGVREVTPEPLNPAKPRRODWDGELYENGSTFYFAKHLIEMGYLQGGKWHHTKC 240
 Db 181 SEIQGVREVTPEPLNPAKPRRODWDGELYENGSTFYFAKHLIEMGYLQGGKWHHTKC 240
 QY 241 ELESVDLDVDIWPFAQVLRVYFGKEKLEKLVLCNDIGLTHGHVVSDDQKEI 300
 Db 241 ELESVDLDVDIWPFAQVLRVYFGKEKLEKLVLCNDIGLTHGHVVSDDQKEI 300
 QY 301 ISYDVKDAIGISLLKKSIEVRLISERACSKQTLSLKLDCMNEVSDKLAWDWERKE 360
 Db 301 ISYDVKDAIGISLLKKSIEVRLISERACSKQTLSLKLDCMNEVSDKLAWDWERKE 360
 QY 361 WGLCWKEVAYLGNESDECLRVGLSGAPADACSYAQKAVGYICKNGGRGATREFAEH 420
 Db 361 WGLCWKEVAYLGNESDECLRVGLSGAPADACSYAQKAVGYICKNGGRGATREFAEH 420
 QY 421 ICLLMKVNNSCQK 434
 Db 421 ICLLMKVNNSCQK 434

RESULT 2

AA90351
 ID AA90351 standard; protein; 434 AA.

AC AA90351;
 XX

DT 04-DEC-2000 (first entry)
 XX

DE Human glycosylation enzyme clone HMLM34 protein sequence.
 XX

KW Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
 KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
 KW haematopoietic cell deficiency; blood coagulation disorder; asthma;

KW afibrinogenaemia; blood platelet disorder; thrombocytopaenia; neoplasia;
 KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
 KW allergic encephalomyelitis; allergic reaction; organ rejection;
 KW graft-versus-host disease; inflammation; hyperproliferative disorder;
 KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase.
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH Misc-difference 133 /note= "encoded by YAT"
 FT Misc-difference 136 /note= "encoded by GKT"
 FT Misc-difference 147 /note= "encoded by YCA"
 FT Misc-difference 169 /note= "encoded by GKT"

PN WO200052136-A2.

PD 08-SEP-2000.

PF 01-MAR-2000; 2000WO-US005325.

PR 02-MAR-1999; 99US-0122409P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Colatan TA;

XX WPI; 2000-572179/53.

XX N-PSDB; AAA37762.

XX New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
 PT synthetase, sialic acid synthetase and aldolase and nucleic acids
 PT encoding the proteins for treating e.g., immune system disorders,
 PT microbial diseases.

PS Claim 12; Page 106-108; 1159p; English.

XX This sequence represents a human glycosylation enzyme clone of the
 CC invention, designated HMLM34. This protein clone is a CMP sialic acid
 CC synthetase. The sequences are useful as reagents for the differential
 CC identification of the tissues or cell types present in a biological
 CC sample, as immunological probes, for treating a disease or condition
 CC and/or treatment of disorders involving aberrant glycolysis, e.g. cramps,
 CC myoglobinuria, and as tumour marker and/or immunotherapy targets. They
 CC may also be used to differentiate, proliferate and attract cells leading
 CC to the regeneration of tissues, to modulate mammalian characteristics
 CC (e.g. in cosmetic surgery) or mammalian metabolism affecting catabolism,
 CC anabolism processing, utilisation and energy storage, to change a
 CC mammal's mental state by influencing and as a food additive or
 CC preservative. The proteins can be used to assay protein levels in a
 CC sample, as a marker or detector of an immune system disorder, to inhibit
 CC cytokine activity, and as a vaccine. They may further be used to treat
 CC immune system or of haematopoietic cell deficiencies or disorders, blood
 CC coagulation disorders (e.g. afibrinogenaemia), blood platelet disorders
 CC (e.g. thrombocytopaenia), wounds resulting from trauma or surgery,
 CC autoimmune disorders (e.g. Addison's disease, multiple sclerosis,
 CC allergic encephalomyelitis), allergic reactions (e.g. asthma, organ
 CC rejection, graft-versus-host disease, inflammation, hyperproliferative
 CC disorders (e.g. neoplasia, purpura, sarcoidosis), diseases caused by
 CC viruses (e.g. hepatitis, meningitis, AIDS), bacteria and fungi (include
 CC e.g. tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia,
 CC cellulitis), and diseases caused by parasites (e.g. amoebiasis,
 CC coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis)

XX Sequence 434 AA;

Query Match 99.6%; Score 2258; DB 3; Length 434;
 Best Local Similarity 100.0%; Pred. No. 2.2e-223;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MDSVEKGAATSVSNPRGPPKIQRNSRGQGRCVKEKPPHIAALIILARGSGKGIPL	60
Db	1	MDSVEKGAATSVSNPRGPPKIQRNSRGQGRCVKEKPPHIAALIILARGSGKGIPL	60
Qy	61	KNIKHLGAVPLIGWVLRALDSGAFQSVWVSTDDHEIENVAKQFGCAQVHRSSSEVSKDSS	120
Db	61	KNIKHLGAVPLIGWVLRALDSGAFQSVWVSTDDHEIENVAKQFGCAQVHRSSSEVSKDSS	120
Qy	121	TSLDAIIIEFLNYXNEXDIVNGIQTATXCLHFTDLOKVAEMIREEGYDSXFSVVRHOFRW	180
Db	121	TSLDAIIIEFLNYXNEXDIVNGIQTATXCLHFTDLOKVAEMIREEGYDSXFSVVRHOFRW	180
Qy	181	SEIQGVREVTPEPLNIPAKPRQDMDGELYENGSFYFAKRHLIEMGYLQGGKWHHTKC	240
Db	181	SEIQGVREVTPEPLNIPAKPRQDMDGELYENGSFYFAKRHLIEMGYLQGGKWHHTKC	240
Qy	241	ELEHSVDIDVDIDWPIAQRVLRVGYFGKELKEIKLIVCNIDGCLTNGHIVYSGDQKEI	300
Db	241	ELEHSVDIDVDIDWPIAQRVLRVGYFGKELKEIKLIVCNIDGCLTNGHIVYSGDQKEI	300
Qy	301	ISYDVKDAIGISLLKKGIEVRLISERACSKOTLSSLKLDCKMEVSVSDKLVAVDEWKE	360
Db	301	ISYDVKDAIGISLLKKGIEVRLISERACSKOTLSSLKLDCKMEVSVSDKLVAVDEWKE	360
Qy	361	MGLCWKEVAVLGNESVDEECLRKRYGLSGAPADACSYACKANGYICKNGGRCATREFAEH	420
Db	361	MGLCWKEVAVLGNESVDEECLRKRYGLSGAPADACSYACKANGYICKNGGRCATREFAEH	420
Qy	421	ICULMEKVNNSCQK	434
Db	421	ICULMEKVNNSCQK	434

RESULT 3

RESOLUTION 3
AAB84682
ID AAB84682 standard: protein; 434 AA.

17-SEP-2001 (first entry)

XX
DE
DE
XX

XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
XX cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
KW vaccine.

OS Homo sapiens.

XX	Key	Location/Qualifiers
FH		

FT Misc-difference 133

```
FT      /note= "unspecified residue encoded by YAT"
```

FT Misc-difference 136

```
FT      /note= "unspecified residue encoded by GKT"
```

AA
PN
WO200142492-A1.

14-JUN-2001.

XX PF 07-DEC-2003: 2000WO-US033136.

XX 09-DEC-1999: 99US-0169839P.

XX
PA (HITMA-) HITMAN GENOME SGT TNC

PA (HUMA-) HUMAN GENOME SCI IN
PA (UYJO) UNIV JOHNS HOPKINS.

PA (UIJO) UNIV JOHNS HO
PA (UTEM) UNIV TEMPLE.

PA (UJEM) UNIV IEMPLE.
PA (UYWY-) JNIV WYCMING.

XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Falter K;
PI Jarvis D;

DR WPI; 2001-441575/47.
DR N-PSDB; AAH28457.

Cells producing cytidine monophosphate-sialic acid and sialylated glycoprotein above endogenous levels for production of vaccines and therapeutics.

PS Claim 20; Fig 30; 182pp; English.

The specification describes a method for manipulating carbohydrate processing pathways in cells of interest. The methods are used to manipulate multiple pathways involved with the sialylation reaction by using recombinant DNA technology and substrate feeding approaches to enable the production of sialylated glycoproteins in the cells. The sialylation process involves the post-translational addition of the donor substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific acceptor carbohydrate. The cells express at least one enzyme, sialylase, from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase, CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The cells are useful for producing complex sialylated glycoproteins in cells of interest, especially insect cells. Glycoproteins containing sialylated oligosaccharides are useful as vaccines, therapeutics and diagnostic tools. Cells producing complex sialylated glycoproteins are useful for enhancing the value of heterologous expression systems and increasing the application of heterologous cell expression products as vaccines, therapeutics and diagnostic tools as well as increasing the variety of heterologous proteins that can be produced and lowering biotechnology production costs. The present sequence represents a human CMP-SA synthetase, which is used in the method of the invention.

Sequence 434 AA:

Query Match

Best Local Similarity 100.0%; Pred. No. 2.2e-223;

Matches	434;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	----	--------	----	------	----

1 MDSVEKGAATSVSNRGRPSRGRPPKLRNSRGGQGRGVEKPFLLAALILARGGSKGIPL 60 QY

Db 1 MDSVEKGAATSVSNPGRGRPSRGRPPXLORNSRGGGGRGVEKPPETAAALILARGGSKGIPL 60

61 KNIKHLAGVPLIGWVLRALDSGAFOSVWVSTDHDEIENVAKOFGAOVHRRSSSEVSKDSS 120

61 KN1KHLACVPLICWTLBAP1DSCREOSVWVSTDDHDEIENAKOECQOVHPPESSEVSKDSS 120

121 TSIDATTEELNYXNEXDIVGNICATSXCIHPTDI.OKVAEMTREETGYDSXESVVRHOFRW 183

121 TS L D A I F E I N Y K N E X D I V C N I Q A T S Y C I H P T D I C K V Z E M I R E E C Y D S Y S V U B P H O E R K 180

181 SETOKCIBZBIMTBDPNI NIDAKBDBZBODWDCBI YENCSSEYBBAKPHI YEMCVI OCCKWEHTTKS 245

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

QY 301 ISYDVNDAIGISLLAKNSGIEVRLTISEKACENQILSSDKNLDCKNEVSVSDKRLAVVDEMRKE 360

DD 301 ISBDVNDATIGISLHUKASGIEVRLHISEKACSRQIESSEHLLDCNIEVSVSDNLAVVDEMKNK 360

361 MGICWKEVAYLGNEVSDDEECLKRVGLSGAPADACSYAKAVGYICKNCNGGRAIRFEAEH 420
QY

DB 361 MGLCWKEVAYLGNVEVSDEECTKRVLGSGAPADACSYAQKAVGYICKNGRGRAIREFAEH 420

QY 421 ICLLMEKVNNSCQK 434

RESULT 4
AAO26545
ID AAO26545 standard; protein; 434 AA.
XX
AC AAO26545;
DT 06-MAR-2003 (first entry)
XX
DE Human CMP-SA synthetase protein.
XX
KW Vaccine; Glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAC; sialic acid;
KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
KW transporter; sialylated glycoprotein; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 133
FT FT /label= His, Tyr
FT /note= "Amino acid residue is encoded by YAT"
FT Misc-difference 136
FT FT /label= Gly, Val
FT /note= "Amino acid residue is encoded by GKT"
FT Misc-difference 147
FT FT /label= Pro, Ser
FT /note= "Amino acid residue is encoded by YCA"
FT Misc-difference 169
FT FT /label= Gly, Val
FT /note= "Amino acid residue is encoded by GKT"
XX
XX US2002142386-A1.
XX
XX 03-OCT-2002.
XX
XX 16-AUG-2002; 2001US-00930440.
XX
XX 02-MAR-1999; 99US-0122582P.
XX 08-DEC-1999; 99US-0169624P.
XX 25-AUG-2000; 2000US-0227579P.
XX
XX (BSTE/) BETENBAUGH M J.
XX (LAWR/) LAWRENCE S.
XX (LEEY/) LEE Y C.
XX (COLE/) COLEMAN T A.
XX
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
XX
XX W2; 2003-102519/09.
XX N-PSDB; AAL53992.
XX
XX Manipulating glycoprotein production in insect cell, involves enhancing
XX expression of enzymes involved in carbohydrate processing pathway such as
XX N-acetylglucosamine-2 epimerase or sialic acid synthetase.
XX
XX Claim 20; Fig 30; 89pp; English.
XX
XX The invention relates to a novel method for manipulating glycoprotein
XX production in an insect cell comprising enhancing expression of an
XX enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
XX catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, sialic acid
XX synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
XX synthetase or CMP-SA transporter, where the expression of each enzyme is
XX enhanced to above endogenous levels. The novel method is useful for
XX manipulating glycoprotein production in an insect cell. Further methods
XX of the invention are useful for producing sialylated glycoprotein. The
XX sialylated glycoprotein produced by the above mentioned methods are
XX useful as pharmaceutical compositions, vaccines, diagnostics and
XX therapeutics. This sequence represents the human CMP-SA synthetase
XX protein of the invention
XX
XX Sequence 434 AA;

Query Match 99.6%; Score 2258; DB 6; Length 434;
Best Local Similarity 100.0%; Pred. No. 2.2e-223; Indels 0; Gaps 0;
Matches 434; Conservative 0; Mismatches 0;
QY 1 MDSVEKGAATSVNPRGRSPRPPKLRNRRGGQGRGVEKPPHLLAALILARGGSGIPL 60
DB 1 MDSVEKGAATSVNPRGRSPRPPKLRNRRGGQGRGVEKPPHLLAALILARGGSGIPL 60
XX
QY 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTDDHDEIENNAKQFQAQVHRSSSEYKDS 120
DB 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTDDHDEIENNAKQFQAQVHRSSSEYKDS 120
XX
QY 121 TSLDAITIEFLNYXNEXDIVGNIQATSKLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 180
DB 121 TSLDAITIEFLNYXNEXDIVGNIQATSKLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 180
XX
QY 181 SEIQGVREVTPEPLNPAKPRRQDWDGELYENGSPYFAKHLEIMGYLQGGKWHHTKC 240
DB 181 SEIQGVREVTPEPLNPAKPRRQDWDGELYENGSPYFAKHLEIMGYLQGGKWHHTKC 240
XX
QY 241 ELEHSVDIDVDIDWPIAEQVRVRYGFGKEKKEIKLVNIDGCLTNGHIYVSGDQKEI 300
DB 241 ELEHSVDIDVDIDWPIAEQVRVRYGFGKEKKEIKLVNIDGCLTNGHIYVSGDQKEI 300
XX
QY 301 ISYDVKDAIGISLLKKGIEVRLISERACSKQTSSLKLDCKMEVSVSKLAVDWMRKE 360
DB 301 ISYDVKDAIGISLLKKGIEVRLISERACSKQTSSLKLDCKMEVSVSKLAVDWMRKE 360
XX
QY 361 MGLCWKEVAYLGNESVDEECLKRVGLSGHPADACSYAQKAVGYICKNGRGRAIFRAH 420
DB 361 MGLCWKEVAYLGNESVDEECLKRVGLSGHPADACSYAQKAVGYICKNGRGRAIFRAH 420
XX
QY 421 ICLMEKVNNSCQK 434
DB 421 ICLMEKVNNSCQK 434
XX
RESULT 5
AAB28673
ID AAB28673 standard; protein; 434 AA.
XX
AC AAB28673;
XX
DT 13-FEB-2001 (first entry)
XX
DE Human carbohydrate-modifying enzyme Incyte ID No: 000422CD1.
XX
KW Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW antiasthmatic; antiarteriosclerotic; antithyroid; hepatotropic;
KW nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW antiarthritic; antiposoriatic; uropathic; ophthalmological;
KW dermatological; antitumor; cytosstatic; virucide; antibacterial;
KW fungicide; protozoacide; tranquilliser; vulnerary; diabetes;
KW autoimmune disorder; inflammatory disorder; infection.
XX
OS Homo sapiens.
XX
XX WO2000063351-A2.
XX
XX 26-OCT-2000.
XX
XX 20-APR-2000; 2000WO-US010882.
XX
XX 21-APR-1999; 99US-0130383P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;
XX WPI; 2000-672723/65.
XX N-PSDB; AAC65392.
XX

PT Novel carbohydrate modifying enzyme polypeptides and polynucleotides for
 PT diagnosis, treatment, and prevention of carbohydrate metabolism
 PT disorders, autoimmune/inflammatory disorders, and cancer.

XX Claim 1; Page 66-67; 75pp; English.

XX The present sequence is a human carbohydrate-modifying enzyme (CME). CME
 CC polynucleotides and polypeptides are useful for treating and diagnosing
 CC diseases associated with CME such as diabetes, autoimmune/inflammatory
 CC disorders such as AIDS, Addison's disease, adult respiratory distress
 CC syndrome, allergies, anaemia, asthma, atherosclerosis, autoimmune
 CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's
 CC disease, emphysema, erythroblastosis fetalis, glomerulonephritis, Good
 CC pasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,
 CC multiple sclerosis, myasthenia gravis, osteoarthritis, osteoporosis,
 CC pancreatitis, polymyositis, psoriasis, Reiter's syndrome, arthritis,
 CC scleroderma, Sjogren's syndrome, systemic lupus erythematosus, ulcerative
 CC colitis, uveitis, Werner syndrome, complications of cancer,
 CC haemodialysis, and extracorporeal circulation, viral, bacterial, fungal
 CC parasitic, protozoal, and helminthic infections, trauma, or cancer. CME,
 CC or its catalytic or immunogenic fragment, is useful for drug screening

XX Sequence 434 AA;

Query Match 96.6%; Score 2188; DB 3; Length 434;

Best Local Similarity 97.0%; Pred. No. 3.7e-216; Indels 0; Gaps 0;
 Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGGGRGVEKPPPELAAIILARGSGSGIPL 60

Db 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGGGRGVEKPPPELAAIILARGSGSGIPL 60

Qy 61 KNIKHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSSVKQSS 120

Db 61 KNIKHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSSVKQSS 120

Qy 121 TSLDAITFEFLNYXNEXDVIQVNIQATSCXCLHPTDLQKVAEMIREGYSDFSVVRHQRW 180

Db 121 TSLDAITFEFLNYHNEVDIVGNIQATSPCLHPTDLQKVAEMIREGYSDFSVVRHQRW 180

Qy 181 SEIQKGVRETEPLNPAKPRQDNDGELYENGFSYFAKRLIEMGYLQGGKWHHTKC 240

Db 181 SEIQKGVRETEPLNPAKPRQDNDGELYENGFSYFAKRLIEMGYLQGGKWHHTKC 240

Qy 241 ELESVDIDVDIDWPIAEORVIRYGFQKELKIKLVNIDCLTNHGYVSGDQKEI 300

Db 241 RAEHSVDIDVDIDWPIAEORVIRYGFQKELKIKLVNIDCLTNHGYVSGDQKEI 300

Qy 301 ISYDVDAIGISLLKSGIEVRLISERACSKQTISSLKLDCKMEVSDKLAVDWEKKE 360

Db 301 ISYDVDAIGISLLKSGIEVRLISERACSKQTISSLKLDCKMEVSDKLAVDWEKKE 360

Qy 361 MGLCKEVAVLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGSGALREFAEH 420

Db 361 MGLCKEVAVLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGSGALREFAEH 420

Qy 421 ICLMEKVNNSCKQ 434

Db 421 ICLMEKVNNSCKQ 434

RESULT 6

AM39319

ID AM39319 standard; protein; 434 AA.

XX AC AM39319;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 2464.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX peripheral nervous system; neuropathy; central nervous system; CNS;

KW

KW

KW

KW

OS

XX

XX

FN

XX

PD

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 leukaemia.

Homo sapiens.

WC200153312-AL.

26-JUL-2001.

26-DEC-2000; 2000WO-US034263.

23-DEC-1999; 99US-00471275.

21-JAN-2000; 2000US-00488725.

25-APR-2000; 2000US-00522317.

20-JUN-2000; 2000US-00598042.

19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

Zhou P, Goodrich R, Drmanac RT;

WPI; 2001-442253/47.

N-PSDB; AAL58475.

Novel nucleic acids and polypeptides, useful for treating disorders such

as central nervous system injuries.

Example 4; SEQ ID NO 2464; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the

encoded polypeptides (AA38642-AA42213) with neurotropic,

immunosuppressant and cytostatic activity. The polynucleotides are useful

in gene therapy. A composition containing a polypeptide or polynucleotide

of the invention may be used to treat diseases of the peripheral nervous

system, such as peripheral neuropathies, peripheral neuropathy and

localised neuropathies and central nervous system diseases, such as

Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

utilization of the activities such as: immune system suppression,

Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

and thrombolytic activity, cancer diagnosis and therapy, drug screening,

assays for receptor activity, arthritis and inflammation, leukaemias and

C.N.S disorders. Note: The sequence data for this patent did not form

part of the printed specification

Sequence 434 AA;

Query Match 96.6%; Score 2188; DB 4; Length 434;

Best Local Similarity 97.0%; Pred. No. 3.7e-216; Indels 0; Gaps 0;

Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGGGRGVEKPPPELAAIILARGSGSGIPL 60

Db 1 MDSVEKGAATSVSNPRGRPPKLRNRRGGGGRGVEKPPPELAAIILARGSGSGIPL 60

Qy 61 KNIKHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSSVKQSS 120

Db 61 KNIKHLAGVPLIGWVLAALDSGAFQSVVSTHDEIENVAKQGAQVHRSSSVKQSS 120

Qy 121 TSLDAITFEFLNYXNEXDVIQVNIQATSCXCLHPTDLQKVAEMIREGYSDFSVVRHQRW 180

Db 121 TSLDAITFEFLNYHNEVDIVGNIQATSPCLHPTDLQKVAEMIREGYSDFSVVRHQRW 180

Qy 181 SEIQKGVRETEPLNPAKPRQDNDGELYENGFSYFAKRLIEMGYLQGGKWHHTKC 240

Db 181 SEIQKGVRETEPLNPAKPRQDNDGELYENGFSYFAKRLIEMGYLQGGKWHHTKC 240

Db 181 SEIQGVREVTPELNINPAKRRQDWDGELZYENGSGFYFAKPHLIEMGYLQGGKMYAYEM 240
 Qy 241 ELEHSVDIDVIDDPIAQRVLRYGFGKEKKEIKLLVCNIDGCLTNHIVYSGDQKEI 300
 Db 241 RAHESVDIDVIDDPIAQRVLRYGFGKEKKEIKLLVCNIDGCLTNHIVYSGDQKEI 300
 Qy 301 ISYDVKDAIGISLLKKSGIEVLISERACSKOTLSLKDCKMEVSVSDKLAVDWEWKE 360
 Db 301 ISYDVKDAIGISLLKKSGIEVLISERACSKOTLSLKDCKMEVSVSDKLAVDWEWKE 360
 Qy 361 MGLCWKEVAYLGNESVDEECLKRVSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420
 Db 361 MGLCWKEVAYLGNESVDEECLKRVSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420
 Qy 421 ICLLMKVNNSCQK 434
 Db 421 ICLLMKVNNSCQK 434

RESULT 7
 AAB95213
 ID AAB95213 standard; protein; 434 AA.
 XX
 AC
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17325.
 XX
 KW Humat; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 CS Homo sapiens.
 XX
 FN EPI:74617-A2.
 XX
 PS 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248016.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 17325; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB955893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 434 AA;

Query Match 96.6%; Score 2188; DB 4; Length 434;

Best Local Similarity 97.0%; Pred. No. 3.7e-216;

Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MDSVEKGAATSVSNPRGRPSRGRPPKLRNSRGGGGRGVEKPPHLLAALLARGSGKGIPL 60
 Db 1 MDSVEKGAATSVSNPRGRPSRGRPPKLRNSRGGGGRGVEKPPHLLAALLARGSGKGIPL 60
 Qy 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTVDHDEIENVAQFQAVHRRSSEVSKDSS 120
 Db 61 KNIKHLAGVPLIGVLRALDSGAFQSVVSTVDHDEIENVAQFQAVHRRSSEVSKDSS 120
 Qy 121 TSLDAIIEFLNYXNEXDIVGNIQATSCXCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 180
 Db 121 TSLDAIIEFLNYXNEXDIVGNIQATSCXCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 180
 Qy 181 SEIQGVREVTPELNINPAKRRQDWDGELZYENGSGFYFAKPHLIEMGYLQGGKMYAYEM 240
 Db 181 SEIQGVREVTPELNINPAKRRQDWDGELZYENGSGFYFAKPHLIEMGYLQGGKMYAYEM 240
 Qy 241 ELEHSVDIDVIDDPIAQRVLRYGFGKEKKEIKLLVCNIDGCLTNHIVYSGDQKEI 300
 Db 241 RAHESVDIDVIDDPIAQRVLRYGFGKEKKEIKLLVCNIDGCLTNHIVYSGDQKEI 300
 Qy 301 ISYDVKDAIGISLLKKSGIEVLISERACSKOTLSLKDCKMEVSVSDKLAVDWEWKE 360
 Db 301 ISYDVKDAIGISLLKKSGIEVLISERACSKOTLSLKDCKMEVSVSDKLAVDWEWKE 360
 Qy 361 MGLCWKEVAYLGNESVDEECLKRVSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420
 Db 361 MGLCWKEVAYLGNESVDEECLKRVSGAPADACSYAQKAVGYICKNGRGGAIRFAEH 420
 Qy 421 ICLLMKVNNSCQK 434
 Db 421 ICLLMKVNNSCQK 434

RESULT 8

AAB43931

ID AAB43931 standard; protein; 448 AA.

XX

AC AAB43931;

XX

DT 03-FEB-2001 (first entry)

XX

DE Human cancer associated protein sequence SEQ ID NO:1376.

XX

XX Human; cancer associated gene; cancer antigen; detection; cancer;
 diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
 KW antidiabetic; antiaschemic; antirheumatic; antithrombotic; antiviral;
 KW antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiac;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nocotropic;
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
 immune disorder; haematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.

XX Homo sapiens.

XX WO200055350-A1.

PN

XX

PD 21-SEP-2000.
 XX 08-MAR-2000; 2000WO-US005882.
 XX 12-MAR-1999; 99US-0124270P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rose: CA, Ruben SM;
 XX WPI; 2000-587533/55.
 DR N-PSDB; AAC78140.
 XX Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer.
 XX Claim 11; Page 2041-2042; 2352pp; English.
 XX AAC77607 to AAC78448 encode the human cancer associated proteins given in
 CC AAB43398 to AAB44239. The proteins can have activities based on the
 CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vulnarary; immunomodulator;
 CC antidiabetic; antiaesthetic; antirheumatic; antiarthritic;
 CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
 CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention
 XX Sequence 448 AA;
 SQ

Query Match 96.6%; Score 2186; DB 3; Length 448;
 Best Local Similarity 97.0%; Pred. No. 3.9e-216;
 Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGGQGRGVEKPEHLAALLIARGGSKIPL 60
 DB 15 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGGQGRGVEKPEHLAALLIARGGSKIPL 74
 QY 61 KNIKHLAGVPLIGWTLRAALDSGAFQSVVSTVDHDEIENVAKQFGAQRHRSSEVSKDS 120
 DB 75 KNIKHLAGVPLIGWTLRAALDSGAFQSVVSTVDHDEIENVAKQFGAQRHRSSEVSKDS 134
 QY 121 TSDAIEFLNYNEXDIVGNIQATSKLHPTDLQKVAEMIREGYSXSVVRHOFRW 180
 DB 135 TSDAIEFLNYNEXDIVGNIQATSKLHPTDLQKVAEMIREGYSXSVVRHOFRW 194
 QY 181 SETQKGVRETEPLNPAKRRQDMDGELYENGSEYFAKRLHIEMGYLQGGKWHYTK 240
 DB 195 SE-QKGVRETEPLNPAKRRQDMDGELYENGSEYFAKRLHIEMGYLQGGKWHYTK 254
 QY 241 ELBHSVDIDVDIWPPIAEQRLVYGYGKGLKEIKLLVNDICLPLNGHLYVSGDQKEI 300
 DB 255 RAEHSVDIDVDIWPPIAEQRLVYGYGKGLKEIKLLVNDICLPLNGHLYVSGDQKEI 314
 QY 301 ISYDVKDAIGLILKKSGLIEVRLISERACSKQTLSLKLCKWEVSVSKLAWDEWKE 360
 DB 315 ISYDVKDAIGLILKKSGLIEVRLISERACSKQTLSLKLCKWEVSVSKLAWDEWKE 374
 QY 361 MGLCWKEVAYLGNSEVDEELCKVGLSGAPADACSQAQKAVGYICKNGGGAIRPFAEH 420
 DB 375 MGLCWKEVAYLGNSEVDEELCKVGLSGAPADACSQAQKAVGYICKNGGGAIRPFAEH 434

QY 421 ICILMEKVNNSCOK 434
 DB 435 ICILMEKVNNSCOK 448
 RESULT 3
 AAB43155
 ID AAB43165 standard; protein; 434 AA.
 XX
 AC AAB43165;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Hunan ORFX ORF2929 polypeptide sequence SEQ ID NO:5858.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnarary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW actinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX Homo sapiens.
 OS
 XX WO2000058473-A2.
 PN
 XX 05-OCT-2000.
 PD
 XX
 PF 31-MAR-2000; 2000WO-US008521.
 XX
 PR 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX Shimkets RA, Leach M;
 PI
 XX WPI; 2000-602362/57.
 DR
 DR N-PSDB; AAC77374.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 11; Page 5028-5029; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnarary;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and actinaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency

CC	{SCID}, AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX	
SQ	Sequence 434 AA;
	Query Match 96.2%; Score 2181; DB 3; Length 434;
	Best Local Similarity 96.8%; Pred. No. 2e-215;
	Matches 42C; Conservative 1; Mismatches 13; Indels 0; Gaps 0;
Qy	1 MDSVEKGAATSVSPRGPRPDKLQRNKGQGQGVKKPPHLLAAILARGGSKGIPL 60
Dd	1 MDSVEKGAATSVSPRGPRPDKLQRNKGQGQGVKKPPHLLAAILARGGSKGIPL 60
Qy	61 KNIKHLAGVPIIGWLVRAALDSGAFQSVMVSTDHDEIENVAKFQAQVHRSSSEVKDSS 120
Dd	61 KNIKHLAGVPIIGWLVRAALDSGAFQSVMVSTDHDEIENVAKFQAQVHRSSSEVKDSS 120
Qy	121 TSLDAITFLNNXEXDIVGNICATSCXLHPDTLOKVAEMIREEGYDSKESVVRHQFRM 180
Dd	121 TSLDAITFLNNHNEVDIVGNIQATSPCLHF ² DLQKVAEMIREEGYDSVFVVRRHQFRM 180
Qy	181 SETQGVREVTPELNINPAKRPRRDNDGELYENGSFYFAKRHLTIEMGYLOGGKWHTTC 240
Dd	181 SETQGVREVTPELNINPAKRPRRDNDGELYENGSFYFAKRHLTIEMGYLOGGKWAYTEM 240
Qy	241 ELESVIDVDIDPIAPORVLRYGFYFKEKLKEIKLVNCIDGLTNGHIYVSQDQEI 300
Dd	241 RAEHSVIDVDIDPIAPORVLRYGFYFKEKLKEIKLVNCIDGLTNGHIYVSQDQEI 300
Qy	301 ISYDVKDAIGISLLKKSIEVRLISERACSKOTLSLKLDCKMEVSVSDKLAVVDWEKE 360
Dd	301 ISYDVKDAIGISLLKKSIEVRLISERACSKOTLSLKLDCKMEVSVSDKLAVVDWEKE 360
Qy	361 MGLCWKEVAYILGNVSDBECLKRVGLSGAPADACSAAKVGYICKNGRGAIREFAEH 420
Dd	361 MGLCWKEVAYILGNVSDBECLKRVGLSGAPADACSAAKVGYICKNGRGAIREFAEH 420
Qy	421 ICLLMKVNNSCOK 434
Dd	421 ICLLMKVNNSCOK 434
RESULT 10	
AAM41105	
ID	AAM41105 standard; protein; 461 AA.
XX	
AC	AAM41105;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 6036.
XX	
KW	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
OS	Homo sapiens.
XX	
EN	WO200153312-A1.
XX	
PT	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US034263.
XX	
PR	23-DEC-1999; 99US-00471275.
PR	21-JAN-2000; 2000US-00488725.
PR	25-APR-2000; 2000US-00552317.
PR	20-JUN-2000; 2000US-00598042.

RESULT 11
ID ADB82754 standard; protein; 245 AA.
XX ADB82754;
AC ADB82754;
XX
XX
DT 04-DEC-2003 (first entry)
XX
DE Human protein sequence useful for the treatment of cancer (SeqID 1535).
XX human; prostate; cancer; cytostatic; gene therapy; vaccine;
KW immune response.
XX
XX Homo sapiens.
XX
XX WO2003050236-A2.
XX
XX 19-JUN-2003.
PD
PF 04-SEP-2002; 2002WO-US028214.
XX
XX 07-DEC-2001; 2001US-00012697.
XX
XX (CHIR) CHIRON CORP.
FA (HYSE-) HYSEQ INC.
XX
XX Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;
PI Crkvenjakov R, Dickson M, Drmanac S, Labat I, Leshkowitz D, Kita D;
PI Garcia V, Jones LW, Stache-Crain B, Scott EM;
XX
XX W21; 2003-513972/48.
XX
XX New polynucleotides derived from human prostate, useful for modulating
PT immune response to prevent or treat cancer.
XX
XX Claim 18; SEQ ID NO 1535; 188pp; English.
XX
XX This invention relates to novel isolated polynucleotides of human origin,
CC particularly isolated from the human prostate. Specifically, it refers to
CC the diagnostics and therapeutics comprising these novel human
CC polynucleotides, and includes the derived probes, antisense
CC oligonucleotides, and antibodies thereof. The identification of these
CC human prostate genes that can inhibit tumour growth is useful for
CC understanding the progression and nature of complex diseases such as
CC cancer, and hence they are important in the drug discovery process. The
CC present invention describes these polynucleotides and encoded
CC polypeptides as exhibiting cytostatic activity, and through gene therapy
CC and/or vaccines they can be used to modulate the immune response for the
CC prevention or treatment of cancers, particularly of the prostate, but
CC also for breast, lung and colon cancer. This polypeptide sequence is a
CC human protein sequence useful for the treatment of cancer, used in an
CC exemplification of the invention. NOTE: These sequences are not given in
CC the specification but are provided on the WIFO website.
XX
XX
SQ Sequence 245 AA;
Query Match 51.8%; Score 1173; DB 7; Length 245;
Best Local Similarity 95.7%; Pred. No. 6.8e-112;
Matches 225; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 137 DIVGNTQATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFVWSEIQKGVREVTPLNL 196
DB 3 DIVGNTQATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFVWSEIQKGVREVTPLNL 62
QY 197 NPAKRPRRQDDGELYENGSGFYFAKHLIEMGYLOGGKWHITTCLEHSDVDIDWDWI 256
DB 63 NPAKRPRRQDDGELYENGSGFYFAKHLIEMGYLOGGKWHITTCLEHSDVDIDWDWI 122
QY 257 AQORVLRGYFGKELKEIKLLVCNIDGCLTNHGIYVSGDQKEIISYDVKDAIGISLLKK 316
DB 23 AQORVLRGYFGKELKEIKLLVCNIDGCLTNHGIYVSGDQKEIISYDVKDAIGISLLKK 182

QY 317 SGIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAJVVDRKEMGLCWKEVAYL 371
DB 183 SGIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAJVVDRKEMGLCWKEVAYL 237
RESULT 12
ABP64227
ID ABP64227 standard; protein; 202 AA.
XX ABP64227;
AC ABP64227;
XX
DT 04-NOV-2002 (first entry)
XX
DE Human ORF597.
XX
XX Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
KW cancer; cardiovascular disease; allergy; autoimmune disease;
KW wound healing; blood coagulation disorder; inflammatory disorder.
XX
XX Homo sapiens.
XX
XX US2002082206-A1.
XX
XX 27-JUN-2002.
XX
XX 30-MAY-2001; 2001US-00867550.
XX
XX 30-MAY-2000; 2000US-0208427P.
XX
XX (LEAC/) LEACH M D.
PA (MEHR/) MEHRABAN F.
PA (CONL/) CONLEY P B.
PA (TOPP/) TOPPER J N.
PA (LAWD/) LAW D.
XX
XX Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
XX
XX WPI; 2002-626554/67.
XX
XX N-PSDB; ABQ98790.
XX
XX New polypeptide designated ORFX are present in human atherogenic cells
PT and are useful to prevent and treat ORFX-associated disorders including
PT cancer, allergy, wound healing or autoimmune, cardiovascular or
PT inflammatory disease.
XX
XX Claim 10; SEQ ID NO 1194; 78pp; English.
XX
XX The present invention relates to novel human ORFX polypeptides and their
CC coding sequences (ABP64227-ABP64227 and ABQ98790-ABQ98790). The sequences
CC were discovered in human atherogenic cells, in particular in platelets
CC and human umbilical vein endothelial cells (HUVEC) and are expressed in
CC many other tissues as well. Atherogenic cells are cells which have the
CC potential to develop atherosclerotic plaques. The ORFX polypeptides and
CC nucleic acids are useful for treating or preventing a pathological
CC condition associated with an ORFX-associated disorder, e.g. cancer,
CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood
CC coagulation disorders or inflammatory disorders. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the USPTC web site at
CC seqdata.uspto.gov/sequence.html?DocID=2002082206
XX
XX Sequence 202 AA;
Query Match 41.7%; Score 945; DB 5; Length 202;
Best Local Similarity 92.2%; Pred. No. 1.6e-88;
Matches 188; Conservative 1; Mismatches 13; Indels 2; Gaps 1;
QY 1 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGQGRGVEKPPHLLAALILARGSKGIPL 60
DB 1 MDSVEKGAATSVNPRGRPSRGRPPKLRNSRGQGRGVEKPPHLLAALILARGSKGIPL 58

QY 61 KNIKHLAGVPLIGWVLRALDSGAFQ 120
 DB 59 KNIKHLAGVPLIGWVLRALDSGAFQ 118
 QY 121 TSLDAITIEFLNXXEXDIDVGNIQATSCCLHPTDQKVAEMIRESGYDSKFXGVRHOPRW 180
 DB 119 TSLDAITIEFLNXXEXDIDVGNIQATSCCLHPTDQKVAEMIRESGYDSKFXGVRHOPRW 178
 QY 181 SEIOKGVREVTPEMLNPAKPRR 204
 DB 179 SEIOKGVREVTPEMLNPAKPRR 202

RESULT 13
 AAU29738
 ID AAU29738 standard; protein; 144 AA.
 XX AC AAU29738;
 XX DC 18-DEC-2001 (first entry)
 XX DE Novel human secreted protein #229.
 XX KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX OS Homo sapiens.
 XX PN WO200179449-A2.
 XX PD 25-OCT-2001.
 XX PF 16-APR-2001; 2001WO-US008656.
 XX PR 18-APR-2000; 2000US-00552929.
 XX PR 26-JAN-2001; 2001US-00770160.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Tang YT, Liu C, Drmanac RT;
 XX DE WPI; 2001-611725/70.
 XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 XX FS Claim 20; Page 186; 765pp; English.
 XX CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 XX secreted proteins of the invention
 SQ Sequence 144 AA;

Query Match 18.5%; Score 419; DB 4; Length 144;
 Best Local Similarity 95.3%; Pred. No. 1.6e-34;
 Matches 82; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNRGPSSRPKLPQNSRGGGGGRGVKPPHLLAAILARGSSKGIPL 60
 DB 15 MDSVEKGAARLPNRPGRPPSSRPKLPQNSRGGGGGRGVKPPHLLAAILARGSSKGIPL 74
 QY 61 KNIKHLAGVPLIGWVLRALDSGAFQ 86
 DB 75 KNIKHLAGVPLIGWVLRALDSGAFQ 100

RESULT 14
 ABP32676
 ID ABP32676 standard; protein; 74 AA.
 XX AC ABP32676;
 XX DT 08-JUL-2002 (first entry)
 XX DE Human ORF1649 protein, SEQ ID NO:3298.
 XX KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulvitis;
 KW vasotrophic; antipsoriatic; antidiabetic; cytostatic; neutropenic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.
 XX OS Homo sapiens.
 XX PN WO200190366-A2.
 XX PD 25-NOV-2001.
 XX PF 24-MAY-2001; 2001WO-US017076.
 XX PR 24-MAY-2000; 2000US-3206690P.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Leach MD, Shimkets RA;
 XX DR WPI; 2002-106200/14.
 XX DR N-PSDB; ABN76702.
 XX PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation.
 XX FS Claim 10; Page 1083; 2508pp; English.
 XX CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNA encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,

CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antifungal activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases
 XX
 SQ Sequence 74 AA;

Query Match 15.0%; Score 339; DB 5; Length 74;
 Best Local Similarity 86.3%; Pred. No. 9.8e-27;
 Matches 63; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 QY 204 RQDWGELYNGSYFAKRLHLEMGYLGQKWHHTKCELEHSDVDIDWPAEQRVLR 263
 Db 1 RQDWGELYNGSYFAKRLHLEMGYLGQKWHHTKCELEHSDVDIDWPAEQRVLR 60
 QY 264 YGYFGKEKLEIK 276
 Db 61 FGFGKEKLEIK 73

RESULT 15
 AAM16058
 ID AAM16058 standard; protein; 44 AA.
 XX
 AC AAM16058;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #2492 encoded by probe for measuring cervical gene expression.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000670.
 XX
 PP 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chec W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 FT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human cervical epithelial cells.

PS Claim 27; SEQ ID NO 20884; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENP; see AAM16058-AAM128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 44 AA;
 Query Match 10.6%; Score 241; DB 4; Length 44;
 Best Local Similarity 100.0%; Pred. No. 5.4e-17;
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 188 REVTEPLNLNPAKRPFRQDWGELYNGSYFAKRLHLEMGYLG 231
 Db 1 REVTEPLNLNPAKRPFRQDWGELYNGSYFAKRLHLEMGYLG 44

Search completed: May 6, 2004, 09:04:19
 Job time : 67.1848 secs


```

RESULT 4
583087
conserved hypothetical protein PA4458 [imported] - Pseudomonas aeruginosa (strain PA01)
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Dec-2002
C/Accession: B83087
3/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; F
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin
.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A/Reference number: A82950; MUID:2043737; PMID:10984043
B/Accession: B83087

```

Query March	9.2%	Score 209.5;	DB 2;	Length 173;
Best Local Similarity	33.6%;	Pred. No. 6.4e-09;		
Matches 51;	Conservative 31;	Mismatches 69;	Indels 1;	Gaps 1
Qy 271	KLKEIKLLVNCNIDGLTNGHYVSSDQKEIISYEKVAIGISLLKSGTEVRLISERACS	330		
Db 13	RASIRLAIFFVDGVLIDGKIYFLVDSGEFZFTLDGHGINKLTIASGVRTAIITORDTP	72		
Qy 331	KQTLSSLLKLDCK-NEVSVSDKLAVVDEWRKMGCLQWKEVAYLGNVSEDECKRVLGSGA	389		
Db 73	WVERARNLGTCHLYQCGREDKLAVLDELLGELGLGYEQVAYLSDDLPLDLPVRRVGLGMA	132		
Qy 390	PADACSTQAQKAVGYICKNGCGGALREFAEHI	421		
Db 133	VASADPFVROHAGHVTAAARGCGGAAREFCELI	164		

RESULT 5
C81785
conserved hypothetical protein NMA2134 [imported] - Neisseria meningitidis (strain Z249)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 23-Dec-2002
C:Accession: C81785
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Moren
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandran
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: A81775; MUID:20223556; PMID:10761919
A:Accession: C81785
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <PAR>
A:Cross-references: GB:AL162758; GB:AL157859; NID:G7380672; PIDN:CAB85346.1; PID:G73807
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA2134
C:Superfamily: phosphatase

Query Match: 9.0%; Score 205; DB 2; Length 178;
 Best Local Similarity 32.9%; Pred. No. 1.4e-08;
 Matches 51; Conservative 35; Mismatches 63; Indels 6; Gaps 3;

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 23-Dec-2002
C:Accession: H85983
R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 469, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; XUID:21074935; PMID:11206551
A:Accession: H85983
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <STO>
A:Cross-references: GB:AE005174; NID:g12517816; PIDN:AA058332.1; GSPDB:GN00145; UWGP:Z45
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Superfamily: phosphatase

Query Match 8.9%; Score 201.5; DB 2; Length 188;
Best Local Similarity 29.2%; Pred. No. 2.8e-08;
Matches 49; Conservative 42; Mismatches 64; Indels 13; Gaps 2;

QY 271 KLEIKLVNIDGCLTNGHIYVSGDQKEIIISYDVKDAIGISLLKKSIEVRLISERACSKQTL 325
DB 22 KAENIRLLILDVDGLSDGLIYMGNGEKLAFNVRDGYGIRCALTSIEVAITGRKAK 81
QY 326 --ERACSKQTLSSKLDCKMEVSVDKLAVDWEKEMGLCWKEVAYLGNVEVSDECLKRVCL 383
DB 62 LVEDRCATLGIHTL-----YQSQSNKLIATFADLLEKLAIPENVAYGDDLDLWPMVEK 135
QY 384 VLSGAPADACSYAOKAVGYICKNGRGGAIRFAEHICLLMEKVNNS 431
DB 136 VLSVAVADAHPLLPRADYVTRIAGRGAVREVCDLILLAQKIDEA 183

RESULT 12
B64716
conserved hypothetical protein HP1570 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 23-Dec-2002
C:Accession: B64716
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen-
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, E.
Nature 388, 539-547, 1997
A:Authors: Mallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; XUID:97394467; PMID:9252185
A:Accession: B64716
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-154 <TOM>
A:Cross-references: GB:AE000654; GB:AE000511; NID:g2314743; PIDN:AAD08607.1; PID:g231475
C:Superfamily: phosphatase

Query Match 8.8%; Score 200; DB 2; Length 164;
Best Local Similarity 29.2%; Pred. No. 3.1e-08;
Matches 49; Conservative 46; Mismatches 57; Indels 16; Gaps 3;

QY 275 IKLLVNDGCLTNGHIYVSGDQKEIIISYDVKDAIGISLLKKSIEVRLISERACSKQTL 334
DB 2 IKLLLDVDTGTLSDGLYFDENFHEIKAFNVDGLGTLWQKLGKXIAITGR----- 54
QY 335 SSLKLDCKME-----VSVDKLAVDWEKEMGLCWKEVAYLGNVEVSDECLKRVCL 386
DB 55 TSIMVKRMESLGVQFVFNVEKNVIERLKKDLQLSAQEIAACVGDYNDLGMFKACAL 114
QY 387 SGAPADACSYAOKAVGYICKNGRGGAIRFAEHICLLMEKVNNS 434
DB 115 SPAPDAHPLLKSKAYKVLQNSGGKAVREADYL-LTLEGLQCEALX 161

RESULT 11
B64716
conserved hypothetical protein HP1570 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 23-Dec-2002
C:Accession: B64716
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen-
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, E.
Nature 388, 539-547, 1997
A:Authors: Mallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; XUID:97394467; PMID:9252185
A:Accession: B64716
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-154 <TOM>
A:Cross-references: GB:AE000654; GB:AE000511; NID:g2314743; PIDN:AAD08607.1; PID:g231475
C:Superfamily: phosphatase

Query Match 8.8%; Score 199.5; DB 2; Length 188;
Best Local Similarity 30.4%; Pred. No. 4e-08;
Matches 51; Conservative 40; Mismatches 64; Indels 13; Gaps 3;

QY 271 KLEIKLVNIDGCLTNGHIYVSGDQKEIIISYDVKDAIGISLLKKSIEVRLISERACSKQTL 325
DB 22 KAENIRLLILDVDGLSDGLIYMGNGEKLAFNVRDGYGIRCALTSIEVAITGRKAK 81
QY 326 --ERACSKQTLSSKLDCKMEVSVDKLAVDWEKEMGLCWKEVAYLGNVEVSDECLKRVCL 383
DB 62 LVEDRCATLGIHTL-----YQSQSNKLIATFADLLEKLAIPENVAYGDDLDLWPMVEK 135

G71802
hypothetical protein jhp1478 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Dec-2002
C:Accession: G71802
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
A:Reference number: A71800; XUID:99120557; PMID:9923682
A:Accession: G71802
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <ARN>
A:Cross-references: GB:AE001569; GB:AE001439; NID:g4156095; PIDN:AAD07050.1; PID:g41560
A:Experimental source: strain J99
C:Genetics:
C:Gene: jhp1478
C:Superfamily: phosphatase

Query Match 8.8%; Score 193.5; DB 2; Length 164;
Best Local Similarity 30.3%; Pred. No. 3.4e-08;
Matches 47; Conservative 43; Mismatches 50; Indels 15; Gaps 2;

QY 275 IKLLVNDGCLTNGHIYVSGDQKEIIISYDVKDAIGISLLKKSIEVRLISERACSKQTL 334
DB 2 IKLLLDVDTGTLSDGLYFDENFHEIKAFNVDGLGTLWQKLGKXIAITGR----- 54
QY 335 SSLKLDCKME-----VSVDKLAVDWEKEMGLCWKEVAYLGNVEVSDECLKRVCL 386
DB 55 TSIMVKRMESLGVQFVFNVEKNVIERLKKDLQLSAQEIAACVGDYNDLGMFKACTL 114
QY 387 SGAPADACSYAOKAVGYICKNGRGGAIRFAEHICLLMEKVNNS 421
DB 115 SPAPDAHPLLKSKAYKVLQNSGGKAVREADYL 149

RESULT 12
AC0905
conserved hypothetical protein SRY3495 [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 23-Dec-2002
C:Accession: AC0905
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
Th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serc
A:Reference number: AB0502; XUID:21534947; PMID:11677608
A:Accession: AC0905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07833.1; PID:g16504381; GSPDB:GX00176
C:Genetics:
C:Gene: SRY3495
C:Superfamily: phosphatase

Query Match 8.8%; Score 199.5; DB 2; Length 188;
Best Local Similarity 30.4%; Pred. No. 4e-08;
Matches 51; Conservative 40; Mismatches 64; Indels 13; Gaps 3;

QY 271 KLEIKLVNIDGCLTNGHIYVSGDQKEIIISYDVKDAIGISLLKKSIEVRLISERACSKQTL 325
DB 22 KAENIRLLILDVDGLSDGLIYMGNGEKLAFNVRDGYGIRCALTSIEVAITGRKAK 81
QY 326 --ERACSKQTLSSKLDCKMEVSVDKLAVDWEKEMGLCWKEVAYLGNVEVSDECLKRVCL 383
DB 62 LVEDRCATLGIHTL-----YQSQSNKLIATFADLLEKLAIPENVAYGDDLDLWPMVEK 135

QY 384 VGLSGAPADACSYAQKAVGYICKNGGGRGATREFAEHICLIMKXNNS 431
 Db 136 VGLSVAVADAPHLIPRADYVTHIAGRGAVREVCDLLLLAQGLDEA 183

RESULT 13
 H61207
 conserved hypothetical protein NMB0353 [imported] - Neisseria meningitidis (strain MC58)
 C:Species: Neisseria meningitidis
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 23-Dec-2002
 C:Accession: H61207
 R:Tectelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, C.; Fleischmann, R.D.; Dougherty, B.A.;
 Li, K.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Eizsa, M.
 Science 237, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter,
 A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; PMID:20175755; PMID:10710307
 A:Accession: H61207
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-178 <TIG>
 A:Cross-references: GB:A8002392; GB:A8002098; NID:G7225573; PIDN:AAE40796.1; PID:G722557
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB0353
 C:Superfamily: phosphatase

Query Match 8.6%; Score 194; DB 2; Length 178;
 Best Local Similarity 31.6%; Pred. No. 1e-07;
 Matches 49; Conservative 36; Mismatches 64; Indels 6; Gaps 3;

QY 274 EIKLLVNCIDCLTNGHYVSGDQKEIISYDVKDAIGISLLKKSIEVRLISERACSKQT 333
 Db 14 KIKLLIDVSDVLDGFRIFRDNGEERKSFETLDGHLKMLQAGVGQTAITEDRPSVG 73

QY 334 LSSLLKLDCKMEV-SVSKLAVVDWRKEMGLCWKEVAYLGNVSDDECLKEVELSGAPAD 392
 Db 74 IRVQLGNIYFKGISDKRAYEELRACAGVEEACAFVGDDVDLPVMVRCGLPVAVPG 133

QY 393 ACSYQAQKAVGYICKNGGGRGATREFAEHIC-LNME 426
 Db 134 AHWFTQHAAYITRHAGGAGVRE---VCDLIWQ 164

RESULT 14
 G64174
 hypothetical protein H1679 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae
 C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 23-Dec-2002
 C:Accession: G64174
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
 Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodex, A.; Kelley, J.M.; Weidman, J.
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; PMID:9530630; PMID:7542800
 A:Accession: G64174
 A:Status: nucleic acid
 A:Molecule type: DNA
 A:Residues: 1-180 <TIG>
 A:Cross-references: GB:U32841; GB:I42023; NID:G1574529; PIDN:AAE23325.1; PID:G1574531; T
 A:Note: best homolog was a hypothetical protein from Escherichia coli
 C:Superfamily: phosphatase

Query Match 8.5%; Score 193.5; DB 2; Length 180;
 Best Local Similarity 29.0%; Pred. No. 1e-07;
 Matches 45; Conservative 41; Mismatches 66; Indels 3; Gaps 2;

QY 269 KEKLKIKLVNCIDCLTNGHYVSGDQKEIISYDVKDAIGISLLKKSIEVRLISER- 327
 Db 2 QQKLENIKFTVDVGLTDGVLTDQLHYDANGBAIKSFHVRDLGLIKMLMDADIQAVVLSGRD 61

QY 328 -ACSKQTLSLLKLDCKMEVSVSKLAVVDWRKEMGLCWKEVAYLGNVSDDECLKRVGL 386
 Db 62 SPILRRRIADLGKLFLLGKLEKETACFD-LMKQAGVTAEQTAYIGDDSDVLPAPACGT 120

QY 387 SGAPADACSYAQKAVGYICKNGGGRGATREFAEHI 421
 Db 121 SFAVADAPYVKNADVHVLSTHGKGRFREMDSMI 155

RESULT 15
 A36509
 N-acylneuraminate cytidyltransferase (EC 2.7.7.43) - Escherichia coli
 C:Species: Escherichia coli
 C>Date: 01-Feb-1991 #sequence_revision 01-Feb-1991 #text_change 08-Oct-1999
 C:Accession: A36509
 R:Zapata, G.; Vann, W.F.; Arcanson, W.; Lewis, M.S.; Moos, M.
 J. Biol. Chem. 264, 14769-14774, 1989
 A:Title: Sequence of the cloned Escherichia coli K1 CMP-N-acetylneuraminic acid synthet.
 A:Reference number: A36509; PMID:89359273; PMID:2549035
 A:Accession: A36509
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-419 <ZAP>
 A:Cross-references: GB:J05023; NID:G146943; PIDN:AAA24210.1; PID:G146944
 C:Keywords: nucleotidyltransferase

Query Match 8.4%; Score 189.5; DB 2; Length 419;
 Best Local Similarity 22.4%; Pred. No. 7e-07;
 Matches 88; Conservative 66; Mismatches 151; Indels 87; Gaps 11;

QY 45 LAAILARAGSGKGPLKNIKHLAGVPLIGWVRALDSGAFQSVVSTHDEIENVAQKF 104
 Db 5 IIAIIPARSGSKGRNKNALMLIDKPLIAYITAEALQSEMFEKVIITVDSQYGAIESY 64

QY 105 GAQVHRSSSVSKDSTSLDAIIBFLVXNEXDVGNTQATXCLHPTDLOKVAEMIRE- 163
 Db 65 GADFLLRPEELATKASFEFKHALSIYDYESFALLQPTSPFPDSTHIEAVKLYOTL 124

QY 164 EGYDSXFSVVRHQRFWSEIQGVREVTPL-----NLNPAKRPRQWWDGELYEN 214
 Db 125 EKYQCVSVVTRSNK-----PSQIRPLDDYSTLSFFDLDYSKYNNS--IVEYHEN 173

QY 215 GSFYFA-KRHLEMGYLOGGKWHYTKCELSHSVDIDVIDMPIA---EQRVLRGYGCKE 270
 Db 174 GAIFIAKQHYLTKHFFGRYSLAYIMDKSSSLDIDRMDPELAIITQQ-----KK 224

QY 271 KLEIKLVNCIDCLTNGHYVSGDQKEIISYDVKDAIGISLL-----KKSGETETLI 324
 Db 225 NRQKIDL-----YQNIENRINEKNEFDSVSDITLIGHSLFDYMWVKKINDIEVNNL 276

QY 325 SERACSKQTLSLLKLDCKMEVSVSKL-----AVVDWRKEMGLCWKEVAYLGNVSG 376
 Db 277 GIAGINSKEYVEYIIEKELIVNFGFVFIFPGTNDIVVSDMKKE----- 320

QY 377 DEBCLKRVGLSGAPADACSYAQKAVGYICKN 408
 Db 321 -----DTLWYLLKTKTCQYIKKN 337

Search completed: May 6, 2004, 09:09:42
 Job time : 17.735 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM : rotein - protein search, using sw model

Run on: May 6, 2004, 08:56:04 ; Search time 11.4127 Seconds
(without alignments)
1380.112 Million cell updates/sec

Title: US-09-930-440B-4

Percent score: 2266

Sequence: 1 MDSVEGAATSVNPRGRPS.....REFAHICLLMEKVNNSQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Sea thed: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result ID	Score	Query Match %	Length	DB ID	Description
1	217	9.6	228	1 NEUA_NEIMB	Q57385 neisseria m
2	210.5	9.3	187	1 KDSC_YERPE	Q82b47 versinia pe
3	202.5	8.9	413	1 NEUA_STR3	Q31598 streptococ
4	201.5	8.9	188	1 KDSC_ECOL6	Q8fd72 escherichia
5	201.5	8.9	188	1 KDSC_ECOLI	P45396 escherichia
6	199.5	8.8	188	1 KDSC_SALTI	Q8z3g5 salmonella
7	199.5	8.8	188	1 KDSC_SALTY	Q8zls0 salmonella
8	198.5	8.8	413	1 NEUA_STR3	Q9afg9 streptococ
9	193.5	8.5	180	1 KDOP_HAEIN	P45314 haemophilus
10	189.5	8.4	419	1 NEUA_ECOLI	P32266 escherichia
11	185.5	8.2	228	1 NEUA_HAEIN	Q57140 haemophilus
12	171	7.5	235	1 PTMB_CAMCO	Q45982 campylobact
13	163.5	7.2	258	1 KDSE_PASMU	P57883 pasteurella
14	131.5	5.8	245	1 KSU5_ECOLI	P42216 escherichia
15	131	5.8	253	1 KDSE_HAEIN	P44490 haemophilus
16	124	5.5	393	1 GLGC_CLOPE	Q8xp97 clostridium
17	121.5	5.4	231	1 KDSE_VIBVU	Q8da99 vibrio vuln
18	119.5	5.3	247	1 KDSE_ECOLI	P04951 escherichia
19	119.5	5.3	250	1 KDSE_YERPE	Q8z9g4 versinia pe
20	118.5	5.2	247	1 KDSE_ECO57	Q8xdg6 escherichia
21	116.5	5.1	247	1 KDSE_SALTI	Q8z800 salmonella
22	116.5	5.1	247	1 KDSE_SALTY	Q8zqco salmonella
23	114.5	5.1	247	1 KDSE_ECOL6	Q8fjd9 escherichia
24	113.5	5.0	234	1 KDSE_AQUAE	Q66914 aquifex aeo
25	110.5	4.9	2194	1 GLSN_MEDSA	Q03460 medicago sa
26	109	4.8	251	1 KDSE_VIBPA	Q87x14 vibrio para
27	108.5	4.8	491	1 PVCA_METTH	Q27939 methanobact
28	107.5	4.7	502	1 LASI_YEAST	P36146 saccharomyc
29	106.5	4.7	722	1 LPSI_CARBL	P90795 caenorhabdi
30	106	4.7	634	1 ICFG_SNNY3	P37979 synchocyst
31	105.5	4.7	246	1 KDSE_RICCN	Q92196 rickettsia
32	105	4.6	475	1 NIFB_ANAAZ	Q43883 anabaena az
33	104	4.6	245	1 KDSE_FUSNN	Q8rfa8 fusobacteri

34	101.5	4.5	252	1 KDSE_VIBCH	Q9kpx2 vibrio chol
35	101.5	4.5	760	1 METE_VIBHA	Q8kr96 vibrio harv
36	99.5	4.4	254	1 KDSE_PSRPK	Q881m7 pseudomonas
37	99	4.4	475	1 NIFB_ANASP	P20627 anabaena sp
38	96	4.2	254	1 KDSE_PSRSM	Q87yf7 pseudomonas
39	96	4.2	293	1 ERA_MYCPU	Q98q11 mycoplasma
40	95.5	4.2	246	1 KDSE_RICHR	Q9zdf0 rickettsia
41	95	4.2	72	1 KSUI_ECOLI	P42215 escherichia
42	94.5	4.2	1076	1 CARB_HALER	Q8r5s3 halomonas e
43	94.5	4.2	3084	1 LMAL_MOUSE	P19137 mus musculu
44	94.5	4.2	8797	1 SNEI_HUMAN	Q8sf91 homo sapien
45	93.5	4.1	340	1 Y483_METJA	Q57907 metatartococ

ALIGNMENTS

RESULT 1
ID NEUA_NEIMB STANDARD; PRT; 228 AA.
AC Q57385;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 13-OCT-2003 (Rel. 42, Last annotation update)
DE Acy-neuraminidase cytidyltransferase (EC 2.7.7.43) (CMP-N-
acetylneuraminic acid synthetase) (CMP-NeuNAC synthetase) (CMP-sialic
acid synthetase).
DE NEUA OR SIAB OR SYNB OR NMB0069.
OS Neisseria meningitidis (serogroup B), and
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491, 487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B1940 / Serogroup B;
RX MEDLINE=93012891; PubMed=1398032;
RA Edwards U., Prosch M.;
RT "Sequence and functional analysis of the cloned Neisseria
meningitidis CMP-NeuNAC synthetase.";
RL FEMS Microbiol. Lett. 75:161-166(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB / Serogroup B;
RX MEDLINE=94156865; PubMed=8113198;
RA Swartley J.S., Stephens D.S.;
RT "Identification of a genetic locus involved in the biosynthesis of N-
acetyl-D-mannosamine, a precursor of the (alpha 2-->8)-linked
RT polysialic acid capsule of serogroup B Neisseria meningitidis.";
RL J. Bacteriol. 176:1530-1534(1994).
RN [3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NTC 8249 / Serogroup B;
RX MEDLINE=94321329; PubMed=8045888;
RA Ganguli S., Zapata G., Wallis T., Reid C., Boulnois G.J., Vann W.F.,
Roberts I.S.;
RT "Molecular cloning and analysis of genes for sialic acid synthesis in
RT Neisseria meningitidis group B and purification of the meningococcal
RT CMP-NeuNAC synthetase enzyme.";
RL J. Bacteriol. 176:4583-4583(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=NRCC 4030 / 406Y / Serogroup Y;
RA Gilbert M., Watson D.C., Wakarchuk W.W.;
RT "Purification and characterization of the recombinant CMP-sialic acid
RT synthetase from Neisseria meningitidis.";
RL Biotechnol. Lett. 19:417-420(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.G.,


```

RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mascen T., Ciecko A., Parksey D.S., Blair S., Clifton H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzia M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -!- CATALYTIC ACTIVITY: CTP + N-acetylneuraminate = diphosphate + CMP-N-
CC acylneuraminate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M95053; AAA20476.1; -
DR EMBL: U43288; AAA17655.1; -
DR EMBL: X78068; AAA54983.1; -
DR EMBL: U60146; AAB60780.1; -
DR EMBL: AEC02366; AAF40536.1; -
DR PIR: B53384;
DR PDB: 1EYR; 14-FEB-01.
DR PDB: 1E2I; 14-FEB-01.
DR TIGR: NMB0069; -
DR InterPro: IPR003329; Cytidylyl trans.
DR Pfam: PF02348; CTP trans 3; 1.
KW Transferase; Nucleotidytransferase; Sialic acid; Complete proteome;
KW 3D-structure.
SQ SEQUENCE 228 AA; 24892 MW; F2510733BFC3A31 CRC64;

Query Match 9.6%; Score 217; DB 1; Length 228;
Best Local Similarity 27.2%; Pred. No. 5.6e-10;
Matches 65; Conservative 40; Mismatches 82; Indels 52; Gaps 6;

QY 47 ALLIARGSKGIPLKNVXHLGAVPLGVLRALDSGAFQSVVWSTDHDETNVAKQFGA 106
D 7 AVILARQNSKGLPKNLKNXNGISLLGHTINAAISSKGFDRIVSTDGGLIAEAKNFV 66
QY 107 QVRRSSSVKSSLSLDAITFNLVXNEXDITVGNIQATXCLPTDLQKVAEMIRESGY 166
D 67 EVLRLPALASDTASSISGVIAL-----ETIGNSGTVLLQPTSLRTCAHIRE--- 117
QY 167 DSXFSV-----VRRHQRWSEIQGVREYVTEPLNLPKAPRRQ----- 205
D 118 -AFSLFDEKIGSVSVSACPMEHPLK-TLLQINNGEVAPNKHLSDLBQPCQLPQAFRP 174
QY 206 -----DWDGELIYENGSYFAKHLIENGYLGGKWHTKCELSHSDVIDWDPIAE 258
D 175 NGAIYINDTASLIANNCFIAPTFLKLYIMSH-----QDSIDIDTLELDLQQA 220

RESULT 2
KDSC VERPE
ID_KDSC_VERPE STANDARD; PRT; 187 AA.
AC Q82B47;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-Deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
DE (KDO 8-P phosphate).
GN KDSC OR YPO3578 OR Y0150.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
CX NCBI_TaxID=632;
RN [1]

```

```

SEQUENCE FROM N.A.
STRAIN-CO-92 / Biovar Orientalis;
MEDLINE=21470413; PubMed=11586360;
Farkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
Farkhill J., Sebailia M., James K.D., Churcher C., Mungall K.L.,
Baker S., Basham D., Bentley S.D., Brooks K., Cordero-Tarraga A.M.,
Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
Leather S., Moute S., Oyston P.C.F., Quail M.A., Rutherford K.,
Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RN Nature 413:523-527(2001).
[2]
SEQUENCE FROM N.A.
STRAIN-KIMS / Biovar Mediaevalis;
MEDLINE=22137853; PubMed=12142430;
Deng W., Burlard V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic
CC phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate +
CC H(2)O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third
CC step.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AJ414157; CAC92807.1; -
DR EMBL: AE013615; AAM83744.1; -
DR InterPro: IPR006549; HAD-SF-IIIa.
DR InterPro: IPR005834; Hydrolase.
DR InterPro: IPR008230; Sugar Phase.
DR Pfam: PF00702; Hydrolase; 1.
DR PIRSF: PIRSF006118; Sugar Phase; 1.
DR TIGRFAMs: TIGR01662; HAD-SF-IIIa; 1.
KW Lipopolysaccharide biosynthesis; Hydrolase; Magnesium;
KW Complete proteome.
SQ SEQUENCE 187 AA; 20299 MW; 295EA75C7DEA7878 CRC64;

Query Match 9.3%; Score 210.5; DB 1; Length 187;
Best Local Similarity 29.1%; Pred. No. 1.4e-09;
Matches 52; Conservative 43; Mismatches 73; Indels 11; Gaps 3;

QY 255 PIAEQRLRYGVFGKELKEIKLVNCDGLTNGHYIVSGDQKELISVDKDAIGISLL 314
D 13 PVADDVI-----QRAAIRLLICDVGWSDGLIYNGQGEKAPNVDGYGIRCL 64
QY 315 KKSGLVRLISERACS--KQTLSIKLDCRMEVSDKLAVVDKRWKMGKLCWEVAVIG 372
D 65 ITSDIDVAITIGRAKULEDRANTLGI-THLYQQSDKLVAHYHELLATLQCOPEQVAVIG 123
QY 373 NEVSDDECLKRVGLSGAPADACSYAKVGVYCKNGGKGAIRFPAEHCILMEKVNNS 431
D 124 DDLIDWPWMAQVGLSVAVADAPHLIPKHYVTRIKGKGAVREVCDLILLAQDKLEGA 182

RESULT 3
NEUA STRA3
ID NEUA_STRA3 STANDARD; PRT; 413 AA.

```

Q53538; Q98055;
01-NOV-1997 (Rel. 35, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Acylneur. aminate cytidyltransferase [EC 2.7.7.43] (CMP-N-acetylneuraminic acid synthetase) (CMP-NeuNAc synthetase) (CMP-sialic acid synthetase).
NECA OR GBS1233.
OS Streptococcus agalactiae (serotype III), and
OC Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=216495, 1311;
[1]
SEQUENCE FROM N.A.
STRAIN=COH1 / Serotype III;
MEDLINE=96228704; PubMed=8810245;
Haft R.F., Wessels M.R., Mebane K.F., Conaty N., Rubens C.E.;
RA "Characterization of cpsf and its product CMP-N-acetylneuraminic acid
RT synthetase, a group B streptococcal enzyme that can function in KI
RT capsular polysaccharide biosynthesis in *Escherichia coli*.";
Mol. Microbiol. 19:555-563(1996).
[2]
SEQUENCE FROM N.A.
STRAIN=COH1 / Serotype III;
MEDLINE=2072630; PubMed=10913080;
Chaffin D.O., Beres S.B., Yim H.H., Rubens C.E.;
RA "The serotype of type Ia and III group B streptococci is determined by
RT the polymerase gene within the polycistronic capsule operon.";
J. Bacteriol. 182:4466-4477(2000).
[3]
SEQUENCE FROM N.A.
STRAIN=NEM316 / Serotype III;
MEDLINE=2242508; PubMed=10464185;
Glaser P., Rusniok C., Buchrieser C., Chevallier F., Frangeul L.,
RA Meadek P., Zouine M., Couve E., Lalioui L., Foyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of *Streptococcus agalactiae*, a pathogen causing
RT invasive neonatal disease.";
Mol. Microbiol. 45:1499-1513(2002).
[4]
SEQUENCE FROM N.A.
STRAIN=O11 / Serotype Ia;
MEDLINE=99395021; PubMed=10464185;
Yamamoto S., Miyake K., Koike Y., Watanabe M., Machida Y., Ohta M.,
RA Iijima S.;
RT "Molecular characterization of type-specific capsular polysaccharide
RT biosynthesis genes of *Streptococcus agalactiae* type Ia.";
J. Bacteriol. 181:5176-5184(1999).
CC -!- CATALYTIC ACTIVITY: CTP + N-acetylneuraminic acid = dihydroxyacetone phosphate + CMP-N-acetylneuraminic acid.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC -!- CAUTION: Has originally (Ref.1) called cpsf.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; U19899; AAB50271.1; -;
DR EMBL; AF163833; AAD53077.1; -;
DR EXBL; AL768849; CAD46892.1; -;
DR EXBL; AB028896; BAA82290.1; -;
DR F.R; S70912; S70912.
DR Sagalistic; gbs1233; -;
DR InterPro; IP003329; Cytidylyl trans.
DR Pfam; PF02348; CTP_transf_3; 1.
DR Exopolysaccharide synthesis; Transferase; Nucleotidyltransferase;
KW Bacterial capsule; Sialic acid; Complete proteome.

FT	CONFLICT	392	413	TQANDLLQSQYQLFVDEVTKL -> DSRG (IN REF. 1 AND 2).	
FT	SEQUENCE	413 AA;	47670 MW;	9089CD673ABD1FC8	CRC64;
QY	Query Match	8.9%; Score 202.5; DB 1; Length 413;			
QY	Best Local Similarity	28.5%; Pred. No. 1.8e-08;			
QY	Matches	79; Conservative	47; Mismatches	112; Indels	39; Gaps
QY	48	LILARGSGKIPKNTKHLAGVPLIGVLRAALDSGAF--QSVWVSTDHDEIENVAQFG	105		
QY	6	IIPARSGSKLPKXMLFLAGKDWIHTIDAAIESGMFDKKDIPVSTDSELYREICL	65		
QY	106	AQVHRSSSVKDS--SLDAIIEPL-NYXNEXDIVGNIQATSCXCLHPTDLQKVAEMIRE	164		
QY	66	ISVWAKPELSTDOA--SYMDKDFLSYEDNQBFV-LLQVTSPLRKSHWIKEMEYYS	124		
QY	165	GYDS--XFSVVRSHQFQWSEI--QKGVREVTPEPLNFEA-KRPREQDWDGLYENGSS	220		
QY	125	DVDNVVVSFSEVEXHPLFTLLSDKGY-----ADMWGADKGYRQDLQPIYYPRG	179		
QY	221	KRLHBMGVYQQGKWHTK-----CELEHSDVID-----VDIDWPIAQR--V	261		
QY	180	NKET----YLRKSFPTSRYAYQMAKEFSLDVDTRDF--HVIGHLFFDVAIREKEN	235		
QY	262	LRGY---FOKEK--KEIKLVACNIDGLTNGHIYVS	295		
QY	236	YKEGYSLEPNREASKIILGDSKTIISLENYHVSQ	272		
RESULT 4					
KOSC	ECOL6				
ID	KOSC	ECOL6	STANDARD;	PRT;	188 AA.
AC	Q8FD72;	Q83JF3;			
DT	10-OCT-2003	(Rel. 42, Created)			
DT	10-OCT-2003	(Rel. 42, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45);				
DE	(KDO 8-P phosphatase).				
GN	KDSC OR C3958 OR SF3238 OR S3456.				
OS	Escherichia coli O6, and				
OS	Shigella flexneri.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Escherichia.				
NCBI	TaxID=217992, 623;				
SEQUENCE FROM N.A.					
SPECIES=B.colli; STRAIN=O6:HL / CFT073 / ATCC 700928;					
MEDLINE=22398234; PubMed=12471157;					
Weich R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,					
Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,					
Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,					
Mobley H.L.T., Donnenberg M.S., Blattner F.R.;					
"Extensive mosaic structure revealed by the complete genome sequence					
of uropathogenic Escherichia coli."					
Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).					
SEQUENCE FROM N.A.					
SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;					
MEDLINE=22272406; PubMed=12384590;					
Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,					
Yang J., Yang X., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,					
Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,					
Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,					
Yu J.;					
"Genome sequence of Shigella flexneri 2a: insights into pathogenicity					
through comparison with genomes of Escherichia coli K12 and O157."					
Nucleic Acids Res. 30:4432-4441(2002).					
SEQUENCE FROM N.A.					
SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;					
MEDLINE=22590274; PubMed=12704152;					
Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,					
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,					

```

RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RA "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RJ Infect. Immun. 71:2775-2786(2003).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic
CC phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 3-phosphate +
CC H2O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third
CC step.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016767; AAN82388.1; -
DR EMBL; AE015335; AAN44704.1; -
DR EMBL; AE016389; AAP18518.1; -
DR InterPro; IPR006549; HAD-SF-IIIa.
DR InterPro; IPR005834; Hydrolase.
DR Pfam; PF00732; Hydrolase; 1.
DR TIGRFAMs; TIGR01662; HAD-SF-IIIa; 1.
DR Lipopolysaccharide biosynthesis; Hydrolase; Magnesium;
KW Complete proteome.
SQ SEQUENCE 188 AA; 20015 MW; 6A40ED8C2CB1E3A7 CRC64;

Query Match 8.9%; Score 201.5; DB 1; Length 188;
Best Local Similarity 29.2%; Pred. No. 7.3e-09;
Matches 49; Conservative 42; Mismatches 64; Indels 13; Gaps 2;

QY 271 KLKEIKLVNIDCLTNGHYVSGDOKEILSYDVKDAIGLSLKKSGIEVRLIS-----325
DB 22 KAENIRLLDVGSLDGLYMGNGEELKAFNVRGIGRCALTSIDIEVAITGRKAK 81

QY 326 --ERACSKQTLSLKLCKMEVSDKLAVVDENKEMGLCKWEKVAYLGNVSDDECLR 383
DB 82 LVEDRCATLGLITL-----YQGSQNKLIASFSLLEKLAIPENVAYVGDLDLDPWVEX 135

QY 384 VLSGADACASYAQKAVGICKNGGRCATREFAEHLCLMKVNN 431
DB 136 VGLSVAVADAPELLIPRADYVTRIAGRGAVREVCDLLLAQGLDEA 183

RESULT 5
KDSC_ECOLI STANDARD; PRT; 188 AA.
AT P45396; P45398;
DC 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
DE (KDO 8-P phosphatase).
GN KDSC OR B3198 OR Z4561 OR ECS4077.
OS Escherichia coli, and
OS Escherichia coli;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562, 83334;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG:1653;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Saiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
[4]
RN CHARACTERIZATION.
RP STRAIN=B;
RX MEDLINE=22625721; PubMed=12639950;
RA Wu J., Woodard R.W.;
RA "Escherichia coli YrBI is 3-deoxy-D-manno-octulosonate 8-phosphate
RT phosphatase.";
RL J. Biol. Chem. 278:18117-18123(2003).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic
CC phosphate.
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate +
CC H2O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third
CC step.
CC -!- SUBUNIT: Homotrimer.
CC -!- MISCELLANEOUS: Optimal pH is 5.5.
CC -!- SIMILARITY: Belongs to the kdsC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18997; AAA57999.1; ALT_FRAME.
DR EMBL; U18997; AAA58000.1; ALT_FRAME.
DR EMBL; AE003399; AAC76230.1; -
DR EMBL; AE005548; AAG58332.1; -
DR EMBL; AP002564; BAB37500.1; -
DR PIR; E91138; E91138.
DR PIR; H65110; H65110.
DR PIR; H85983; H85983.
DR EcoGene; EG12804; kdsC.
DR InterPro; IPR006549; HAD-SF-IIIa.
DR InterPro; IPR005834; Hydrolase.
DR InterPro; IPR008230; Sugar Phase.
DR Pfam; PF00702; Hydrolase; 1.
DR PIRSF; PIRSF006118; Sugar Phase; 1.
DR TIGRFAMs; TIGR01662; HAD-SF-IIIa; 1.
DR Lipopolysaccharide biosynthesis; Hydrolase; Magnesium;
KW Complete proteome.
SQ SEQUENCE 188 AA; 19997 MW; 75AF512CF4E556E2 CRC64;

Query Match 8.9%; Score 201.5; DB 1; Length 188;

```

Best Local Similarity 29.2%; Pred. No. 7.3e-09;
Matches 49; Conservative 42; Mismatches 64; Indels 13; Gaps 2;

QY 271 KLKEIKLLVNCIDGCLTNGHIYVSGQKEIISYDVKDAGISLLKXSGIEVRLIS----- 325
D5 22 KAENIRLLILDVGVLSGLIYMGNGEELKAFNRVDGYGIRCALTSNIEVAITGRKAK 81

QY 326 --ERACKQTLSKLDCKMEVSVSKLAVVDWKRKMGKLCWEKVAYLGNVSEDECKR 383
D5 82 LVEDRCATLGIHTL-----YQGSNKLIAFSLLEKLAIPENVAVVGDDLDWPMVEK 135

QY 384 VGLSGAPADACSVAKQAVGICKNGRGGAIRFAEHICLLMEKVNS 431
D5 136 VGLSVAVAADAHPLIIPRADIYVTHIAGRGAVREVCDDLLLAQGLDEA 183

RESULT 6
KDS SC SALT
ID KDS SC SALT
AC Q8ZJG5; STANDARD; PRT; 188 AA.

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
DE (KDO 8-P phosphatase)
GN KDS OR STY3495 OR T3233.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham P., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krost A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18."
RT Nature 413:848-852(2001).
RE [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18."
RT J. Bacteriol. 185:2330-2337(2003).
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate + H2O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AL627278; CAD07833.1; -.

DR EMBL; AE016845; AA070769.1; -.
DR InterPro; IPR006549; HAD-SF-IIIa.
DR InterPro; IPR005834; Hydrolase.
DR InterPro; IPR008230; Sugar Ptase.
DR Pfam; PF00702; Hydrolase; I.
DR PIRSF; PIRSF006118; Sugar Ptase; 1.
DR TIGRFAMS; TIGRF01662; HAD-SF-IIIa; 1.
KM Lipopolysaccharide biosynthesis; Hydrolase; Magnesium;
KW Complete proteome.
SQ SEQUENCE 188 AA; 20075 MW; ADEFB0323DBFEF43 CRC64;

Query Match 3.8%; Score 199.5; DB 1; Length 188;
Best Local Similarity 30.4%; Pred. No. 1e-08;
Matches 51; Conservative 40; Mismatches 64; Indels 13; Gaps 3;

QY 271 KLKEIKLLVNCIDGCLTNGHIYVSGQKEIISYDVKDAGISLLKXSGIEVRLIS----- 325
D5 22 KAENIRLLILDVGVLSGLIYMGNGEELKAFNRVDGYGIRCALTSNIEVAITGRKAK 81

QY 326 --ERACKQTLSKLDCKMEVSVSKLAVVDWKRKMGKLCWEKVAYLGNVSEDECKR 383
D5 82 LVEDRCATLGIHTL-----YQGSNKLIAFSLLEKLAIPENVAVVGDDLDWPMVEK 135

QY 384 VGLSGAPADACSVAKQAVGICKNGRGGAIRFAEHICLLMEKVNS 431
D5 136 VGLSVAVAADAHPLIIPRADIYVTHIAGRGAVREVCDDLLLAQGLDEA 183

RESULT 7
KDS SC SALT
ID KDS SC SALT
AC Q8ZJG5; STANDARD; PRT; 188 AA.

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)
DE (KDO 8-P phosphatase)
GN KDS OR STM3316.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du R., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan K., Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2."
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic phosphate (By similarity).
CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate + H2O = 3-deoxy-D-manno-octulosonate + phosphate.
CC -!- COFACTOR: Magnesium (Probable).
CC -!- PATHWAY: Lipopolysaccharide biosynthesis; KDO biosynthesis; third step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the kdsC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE008852; AAL22185.1; -.

[2]
 RN FUNCTION, AND COFACTOR.
 RX MEDLINE=22625721; PubMed=12639950;
 RA Wu J., Woodard R.W.;
 RT "Escherichia coli YrbI is 3-deoxy-D-manno-octulosonate 8-phosphate
 RL J. Biol. Chem. 278:18117-18123(2003).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.67 ANGSTROMS), AND METAL BINDING.
 RX MEDLINE=21824378; PubMed=11935514;
 RA Parsons J.P., Lim K., Tempczyk A., Krajewski W., Eisenstein E.,
 RA Herzberg O.;
 RT "From structure to function: YrbI from Haemophilus influenzae
 RL Proteins 46:393-404(2002).
 CC -!- FUNCTION: Catalyzes the hydrolysis of KDO 8-P to KDO and inorganic
 CC phosphate.
 CC -!- CATALYTIC ACTIVITY: 3-deoxy-D-manno-octulosonate 8-phosphate +
 CC H₂O = 3-deoxy-D-manno-octulosonate + phosphate.
 CC -!- COFACTOR: Magnesium (Probable).
 CC -!- MISCELLANEOUS: Cobalt was used in the crystallography experiment
 CC but magnesium is likely to be the physiological metal.
 CC -!- SIMILARITY: Belongs to the kdsC family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U32841; AAC23325.1; -.
 DR PIR; G64174; G64174.
 DR PDB; 1J8D; 12-AUG-03.
 DR PDB; 1KIE; 12-AUG-03.
 DR TIGR; H1679; -.
 DR InterPro; IPR006549; HAD-SF-III.A.
 DR InterPro; IPR005834; Hydrolase.
 DR InterPro; IPR008230; Sugar Phase.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PIRSF; PIRSF06118; Sugar Phase; 1.
 DR TIGSFAMS; TIGSF01662; HAD-SF-III.A; 1.
 DR KW Hydrolase; Metal-binding; Magnesium; 3D-structure; Complete proteome.
 FT METAL 14 14 MAGNESIUM (PROBABLE).
 FT METAL 16 16 MAGNESIUM (PROBABLE).
 FT METAL 107 107 MAGNESIUM (PROBABLE).
 SQ SEQUENCE 180 AA; 19432 MW; 23CD435E4E83A095 CRC64;
 Query Match 8.5%; Score 193.5; DB 1; Length 180;
 Best Local Similarity 29.0%; Pred. No. 2.9e-08;
 Matches 45; Conservative 41; Mismatches 66; Indels 3; Gaps 2;
 QY 269 KEKKEIKLVNIGCLTNGHYVSGDQKEIISYDVDAIGISLILKSGIEVRLISER- 327
 DB 2 OOKLNRIKIEVITDVGVLTDGQHYDANGEARKSFHRLGLKMLMDADIQVAVLSGRD 61
 QY 328 -ACSKQTLSLKDCKMEVSDKLAVDNEKMGKLCWKEVAYLGNVSDDECLKRVGL 386
 DB 62 SPILRRRADIIGKILFFLGKLEKETACFD-LNKQGVTRAEQTAYIGDSVDLPAFAACGT 120
 QY 387 SGAPADACSQAQVAGYICKNGGKGATREFABHI 421
 DB 121 SPVADAPTYVKNVDVHLSVTHGGKGAFFEMSDMI 155
 RESULT 10
 NEUF ECOLI
 ID NEUF ECOLI
 AC P13266;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Acylneuraminate cytidyltransferase (EC 2.7.7.43) (CMP-N-
 GN acetylneuraminic acid synthetase) (CMP-NeuNAc synthetase).
 OS NEUA.
 OC Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 RN NCEI_TaxID=562;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KL;
 RX MEDLINE=89359273; PubMed=2549035;
 RA Zapata G., Vann W.F., Aaronson W., Lewis M.S., Moos M.;
 RT "Sequence of the cloned Escherichia coli Ki CMP-N-acetylneuraminic
 RL acid synthetase gene.";
 RL J. Biol. Chem. 264:14769-14774(1989).
 CC -!- CATALYTIC ACTIVITY: CTP + N-acylneuraminate = diphosphate + CMP-N-
 CC acylneuraminate.
 CC -!- COFACTOR: Magnesium; other divalent cations are less effective.
 CC -!- SUBUNIT: Homodimer (Probable).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J05023; AAA24210.1; -.
 DR PIR; A36509; A36509.
 DR InterPro; IPR003329; Cytidylyl trans.
 DR InterPro; IPR001087; Lipase GDSL.
 DR Pfam; PF02348; CTP trans 3; 1.
 DR Pfam; PF00857; Lipase GDSL; 1.
 KW Transferase; Nucleotidyltransferase; Sialic acid; Magnesium.
 SQ SEQUENCE 419 AA; 48736 MW; AB8E7783CDD74F19 CRC64;
 Query Match 8.4%; Score 189.5; DB 1; Length 419;
 Best Local Similarity 22.4%; Pred. No. 1.9e-07;
 Matches 88; Conservative 66; Mismatches 151; Indels 87; Gaps 11;
 QY 45 LAELILARGSGKIPLNKIKHLAGVPLIGWVLAALDGAQFQSVWVSTHDEIENVAQKF 104
 DB 5 IIAIIPARGSGKLRNKNALMLIDKPLLAYTTEALQSEMFEKVIIVTDTSEQYGAIAESY 64
 QY 105 GQVHRSEVSKDSTSLDAIEFLNLYNEXDIVGNIQATSCXCLPTDLQKVAEMIRE- 163
 DB 65 GADELRLPELATDKASSPEFIKHALSIYTDYESFALLOFTSPFRDSTHIEAVKLYQTL 124
 QY 164 EGYDSKFSVVRHQFWSBIQGVREVEPL-----NLNFAKPRRRQDWDGELYEN 214
 DB 125 ERYQCVSVSTRNK-----PSQIIRPLDDYSTLSFFDLDSKYKNRS--IVYHPN 173
 QY 215 GSFYTA-KRHLEMGYLGKQGWHTTKCELEHSVDIDVDWPIA---EQRVLYRYFGKE 270
 DB 174 GAFLANKQHYLHTKHFGRYSLAYIMDKESLIDIDRMDFELAITQQ-----KK 224
 QY 271 KLEKIKLLVCNIDGLTNGHYVSGDQKEIISYDVDAIGISL-----KKSIEVRLI 324
 DB 225 NFQKIDL-----YQNHNRINERKNEFDSVSDITLGHSLFDYWDVKINDIEVNNL 276
 QY 325 SERAKSKOTLSLKDCKMEVSDKL-----AVDEWRKMGKLCWKEVAYLGNVSD 376
 DB 277 GIAGINSKEYEYIIEKELIVNFGFVFIFFGTNDIVSDMKK----- 320
 QY 377 DEECLKRVLSGAPADACSQAQVAGYICKN 428
 DB 321 -----DTLWYLNKTKQYIKKN 337
 RESULT 11


```

NEUA HAEIN
ID NEUA HAEIN STANDARD; PRT; 228 AA.
AC Q57140; OC5551;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable acylneuraminase cytidyltransferase (EC 2.7.7.43) (CMP-N-
DE acetylneuraminic acid synthetase) (CMP-NeuNAc synthetase) (CMP-sialic
DE acid synthetase).
GN NEUA OR H1279.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
CX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karlvag A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs C., Hedblom B., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.C., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512 (1995).
CC -!- CATALYTIC ACTIVITY: CMP + N-acylneuraminate = diphosphate + CMP-N-
CC acylneuraminate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32807; AAC22927.1; --
CC TIGR; H1279; --
CC InterPro; IPR003329; Cytidylyl_trans.
CC Pfam; PF02348; CTP_transf_3; 1.
CC Transferrase; Nucleotidyltransferase; Sialic acid; Complete proteome.
CC SEQUENCE 228 AA; 25308 MW; B3CC7674CA379551 CRC64;
Query Match 8.2%; Score 185.5; DB 1; Length 228;
Best Local Similarity 27.9%; Pred. No. 1.7e-07;
Matches 65; Conservative 40; Mismatches 9; Indels 35; Gaps 9;
QY 47 ALIARGSGKGIPLKNIKHLGVLIGWVLRALDAGAFOSVWVSTHDEIENVAQFGR 106
DB 9 ALIARGSGKGIPLKNIKHLGVLIGWVLRALDAGAFOSVWVSTHDEIENVAQFGR 68
QY 107 QVHRSSSEVSKDSSTSLDAIEFLNYKXEXDIVGNIGTSCXCLHPT-----DLQKVAE 159
DB 69 KPVARPESAQSDTRTIDAILHCLSTLIS-----QGTAAALQPTSPRLNALDIRNAME 122
QY 160 MIREBGVDSXSV--VRHQFQ-----WSEIQGVREYTEPLNINPAKRPRQDWDGBLY 212
DB 123 IFLGKYKSVSVSACEHPKYSFTLEGTEVQP-IHEITD-FESPRQKPKSYRANGIY 180
QY 213 ENG--SFYAKRHLLIE--MGYLQGGKWHHTKCELSHSDVIDIDWPIAEQV 261
DB 181 INDIQSLFEERKFFAPMRFYLM-----PTVRSIDIDSTLDQLAESLI 224
RESULT 12

```

```

PTMB CAMCO
ID ETWS CAMCO STANDARD; PRT; 235 AA.
AC Q45982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Posttranslational flagellin modification protein B.
GN ETWS.
OS Campylobacter coli.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
CX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VCL167;
RX MEDLINE=96423180; PubMed=8825781;
RA Guerry P., Doig P., Alm R.A., Burr D.H., Kinsella N., Trust T.J.;
RA "Identification and characterization of genes required for post-
RA translational modification of Campylobacter coli VCL167 flagellin.";
RL Mol. Microbiol. 19:369-378 (1996).
CC -!- FUNCTION: REQUIRED FOR BIOSYNTHESIS OF LAH MODIFICATION IN THE
CC POST-TRANSLATIONAL MODIFICATION OF CAMPYLOBACTER COLI FLAGELLIN.
CC -!- SIMILARITY: Belongs to the CMP-NeuNAc synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U25992; AAB48074.1; --
CC PIR; S70685; S70685.
CC InterPro; IPR003329; Cytidylyl_trans.
CC Pfam; PF02348; CTP_transf_3; 1.
CC SEQUENCE 235 AA; 26600 MW; A58421F922DA040B CRC64;
Query Match 7.5%; Score 171; DB 1; Length 235;
Best Local Similarity 26.0%; Pred. No. 2.5e-06;
Matches 63; Conservative 50; Mismatches 95; Indels 30; Gaps 10;
QY 49 ILARGSGKGIPLKNIKHLGVLIGWVLRALDAGAFOSVWVSTHDEIENVAQFQAV 108
DB 8 ICARGSGKGVKNKIRKINDLEMIAYSIIQAKSGKLFKHIVISTDSEIATKLYGGEV 67
QY 109 -HRSSEVSKDSSTSL-----DAIEFLNYXN-EXDIVGNIGTSCXCLHPTDLQKVAEMIR 162
DB 68 FFKREAHLASATAKIPVWRQALLRSEYFFKQCFDILLDASAPLRSSADIIKAFETFC 127
QY 163 BEGYDSXFSVV--RRHQFRWSEIQGVREYTEPLNINPAKR---PREQDWDGELYENG 216
DB 128 QNQNDLITAVPARNPYFNLEIVQDG--KVVKSGNGFTTRQSVPKCYDM-----NAS 179
QY 217 FYFAGRHLEMGYLOGGKWHHTKCELSHV--DIDVIDWPIAEQVRLRYGVFGKELKE 274
DB 180 IYIFKRDPELLQDCSVFGK-NTGLFIMDESTAFDVSSELDKFIVE-----FLIKEKNLQ 231
QY 275 IK 276
DB 232 AK 233
RESULT 13
KDSB PASMU
ID KDSB PASMU STANDARD; PRT; 258 AA.
AC P57883;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-manno-oculosonate cytidyltransferase (EC 2.7.7.38) (CMP-KDO
DE synthetase) (CMP-2-keto-3-deoxyoctulosonic acid synthetase) (CKS).
GN KDSB OR PWC856.

```


KDSB HAETK
ID KDSB HAETK STANDARD; PRT; 253 AA.
AC P44490;
DI 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-deoxy-manno-oculosonate cytidyltransferase (EC 2.7.7.38) (CMP-KDO
synthetase) (CMP-2-keto-3-deoxyoculosonic acid synthetase) (CKS).
GN KDSB OR H10058.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI TaxID=727;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350830; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fife L.D., Fritchman J.H., Fuhrman J.L., Geoghagen N.S.M.,
RA Gneam C.L., McDonald L.A., Smal K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd";
RL Science 269:496-512(1995).
CC -!- FUNCTION: Activates KDO (a required 8-carbon sugar) for
CC incorporation into bacterial lipopolysaccharide in Gram-negative
CC bacteria (By similarity).
CC -!- CATALYTIC ACTIVITY: CTP + 3-deoxy-D-manno-oculosonate =
CC diphosphate + CMP-3-deoxy-D-manno-oculosonate.
CC -!- PATHWAY: Lipopolysaccharide biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the kdsB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32691; AAC21736.1; -.
CC F1R; G64045; G64045.
CC T1R; H10058; -.
CC HAXAP; MF 00057; -; 1.
CC InterPro; IPR003329; Cytidylyl_trans.
CC InterPro; IPR004528; KdsB.
CC Pfam; PF02348; CTP_transf_3; 1.
CC TIGRfams; TIGR00466; KdsB; 1.
CC Kw Lipopolysaccharide biosynthesis; Transferase; Nucleotidyltransferase;
CC Complete proteome.
CC INIT MET 0 BY SIMILARITY.
CC SEQUENCE 253 AA; 28124 MW; 2F95E2B15BCB56A2 CRC64;
Query Match 5.5%; Score 131; DB 1; Length 253;
Best Local Similarity 28.6%; Pred. No. 0.0041;
Matches 50; Conservative 32; Mismatches 69; Indels 24; Gaps 7;
QY 48 LILARGGSGKPLKNIKHAGVPLIGWVLRALDSGAFQSVWVSTDDHDEIENNAKQPCNAQ 107
DB 5 LIPARFASRLPGKPLADIKGKPMIQVTFERKALQSGA-SRVIIATDNNVADVAKSFGAE 63
QY 108 VHRASSEVSKDSSTG-LDAITIEFLNXXNEXDIVGNQATSCXCLHPTDLQKVAEMIREEGY 166
DB 64 VCMTS--VYHNSGTIERLAEVVEKL-AIPDNEIIVWICGDEPLIPPIVVRQVADNLAKENV 120
QY 167 DSXFSVTRHPRFRSEIOKGVREYVEPLNLPAPKPRQDWDGELYNGSYFAK 221

Db 121 JMASLAVKIH-----DAELENFNAVK--VLTKDGYV-----LYFSR 156

Search completed: May 6, 2004, 09:05:00
Job time : 13.4127 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Rur on: May 6, 2004, 08:59:19 ; Search time 42.4806 Seconds
(without alignments)
3223.466 Million cell updates/sec

Title: US-09-930-440B-4

Perfect score: 2266

Sequence: 1 MDSVEKGAATSVNPRGRPS.....REFAEHICILMEKVNNSCQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	2198	97.0	434	4 Q8NFW8	Q8nfw8 homo sapien
2	2188	96.6	434	4 Q9NQZ0	Q9nqz0 homo sapien
3	2052	90.6	432	11 Q8K2G7	Q8k2g7 mus musculus
4	2048	90.4	432	11 Q88719	Q88719 mus musculus
5	2299	57.3	263	4 Q96AX5	Q96ax5 homo sapien
6	1107.5	48.9	432	13 Q90WG6	Q90wg6 oncofrynchu
7	952	42.0	200	11 Q8C330	Q8c330 mus musculus
8	813	35.9	166	11 Q99KK2	Q99kk2 mus musculus
9	552	24.4	330	17 Q8TJL6	Q8tjl6 methanocarc
10	402	17.7	424	16 Q9AK46	Q9ak46 streptomyce
11	376	16.6	432	16 Q82HY3	Q82hy3 streptomyce
12	288.5	12.7	234	16 Q8KDA1	Q8kda1 chlorobium
13	268	11.8	187	5 Q8IQV0	Q8iqv0 drosophila
14	265	11.7	228	16 Q8F5R2	Q8f5r2 leptospira
15	250	11.0	229	16 Q89HL8	Q89hl8 bradyrhizob
16	231.5	10.2	179	16 Q87WU4	Q87w4 pseudomonas

17	230.5	10.2	163	16 Q67920	Q67920 aquifex aeo
18	228	10.1	233	16 Q8DE02	Q8de02 vibrio vuln
19	226.5	10.0	227	2 Q8R9M4	Q8r9m4 rhizobium m
20	225.5	10.0	169	2 Q9F2B5	Q9f2b5 thaxtera aro
21	223.5	9.9	174	2 Q8KLM5	Q8klw5 pseudomonas
22	223	9.8	209	16 Q8F5Q2	Q8f5q2 leptospira
23	223	9.8	421	2 Q93NQ1	Q93nq1 escherichia
24	217.5	9.6	164	16 Q8A7L2	Q8a7l2 bacteroides
25	217.5	9.6	188	16 Q7U9R7	Q7u9r7 synechococc
26	211.5	9.3	174	16 Q88P96	Q88p96 pseudomonas
27	209.5	9.2	179	16 Q8HV99	Q8hv99 pseudomonas
28	209	9.2	239	16 Q87T67	Q87t67 vibrio para
29	205.5	9.1	232	2 Q8BDX4	Q8bdx4 legionella
30	205	9.0	178	16 Q8JUS3	Q8jsu3 neisseria m
31	203.5	9.0	183	16 Q8EAE9	Q8eae9 shewarella
32	203.5	9.0	228	2 Q8R9S4	Q8r9s4 aeromonas p
33	203.5	9.0	413	2 Q9ALW4	Q9alw4 streptococc
34	202.5	8.9	229	2 Q8KHC9	Q8khc9 pseudomonas
35	202	8.9	226	16 Q7U960	Q7u960 synechococc
36	201.5	8.9	413	2 Q93T10	Q93t10 streptococc
37	200	8.8	164	16 Q26090	Q26090 helicobacte
38	199.5	8.8	164	16 Q9ZJ39	Q9zj39 helicobacte
39	199	8.8	172	16 Q8KE66	Q8ke66 chlorobium
40	199	8.8	238	16 Q82UC4	Q82uc4 nitrosomona
41	197	8.7	168	16 Q8EGS1	Q8egs1 fusbacteri
42	194	8.6	178	16 Q9KL37	Q9kl37 neisseria m
43	193	8.5	223	16 Q9CP68	Q9cp68 pasteurella
44	192.5	8.5	165	16 Q83DI2	Q83di2 coxiella bu
45	187	8.3	233	16 Q8SLF1	Q8slf1 bradyrhizob

ALIGNMENTS

RESULT 1

Q8NFW8 PRELIMINARY; PRT; 434 AA.

AC Q8NFW8;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Cytidine monophospho-sialic acid synthase.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lawrence S.M., Huddleston K.A., Tomiya N., Nguyen N., Lee Y.C.,
Vann W.F., Coleman T.A., Betenbaugh M.J.,
RT "Cloning and Expression of Human Sialic Acid Pathway Genes to Generate
CMP-Sialic Acids in Insect Cells."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF397212; AAM90580.1; -
DR Genew; HGNC:18290; CMAS.
DR GO; GO:0003103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl_trans.
DR Pfam; PF02348; CTP_transf_3; 1.
SQ SEQUENCE 434 AA; 48449 MW; C8FAD3D8CDCF835B CRC64;

Query Match 97.0%; Score 2198; DB 4; Length 434;
Best Local Similarity 98.2%; Pred. No. 2.4e-173;
Matches 426; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy	1	MDSVEKGAATSVNPRGRPSGRPPKLRNSRGQGQGVGKPPHPLAALILARGSKGIPL 60
Db	1	MDSVEKGAATSVNPRGRPSGRPPKLRNSRGQGQGVGKPPHPLAALILARGSKGIPL 60
Qy	61	KNKHLAGVPLIGWLVRAALDSGAFQSVWSTTHDEIENVAKQFGAQVHRSRSEVSKDS 120
Db	61	KNKHLAGVPLIGWLVRAALDSGAFQSVWSTTHDEIENVAKQFGAQVHRSRSEVSKDS 120
Qy	121	TSLDAIIEFLNXXNEXDIVGNIQATSCNCLHTDLQKVAEMIREEGYDSXFSVVRHQFRW 180

```

Db      121  TSLDAITEFLNYXEXDVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
QY      181  SEIQGVRETEPLNINPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
Db      181  SEIQGVRETEPLNINPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
QY      241  SEIQGVRETEPLNINPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
Db      241  ELESVDIDVDIDPIAEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
QY      241  RAERSVDIDVDIDPIAEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
Db      301  ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVIDEWRKE 360
QY      301  ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVIDEWRKE 360
Db      361  MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSQAQKAVGYICKNGRGGAIRFAEH 420
QY      361  MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSQAQKAVGYICKNGRGGAIRFAEH 420
Db      421  ICLLMKVNNSCQK 434
QY      421  ICLLMKVNNSCQK 434
Db      421  ICLLMKVNNSCQK 434

RESULT 2
Q9NQ20 PRELIMINARY; PRT; 434 AA.
AC Q9NQ20;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE CMP-N-acetylneuraminic acid synthase (EC 2.7.7.43).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bouquin T., Mundy J.;
RT "Human mRNA for CMP-N-acetylneuraminic acid synthase.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Isogai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saico K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y., Oshima A.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF271388; AAF76203.1; -.
DR EMBL; AK022927; BAB1411.1; -.
DR GO; GO:0006781; F:N-acylneuraminate cytidyltransferase acti. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl trans.
DR InterPro; IPR000794; Ketoacyl synth.
DR Pfam; PF02348; CTP transf 3; 1.
DR PROSITE; PS00606; S_KETOACYL_SYNTHASE; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 434 AA; 48379 MW; B303B6647EF81A3A CRC64;

Query Match 96.6%; Score 2188; DB 4; Length 434;
Best Local Similarity 97.0%; Pred. No. 1.6e-172;
Matches 421; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGSRGPPKLCNSRGGGQGRGVEKPPHLLAAILARGSGKGIPL 60
Db 1 MDSVEKGAATSVSNPRGSRGPPKLCNSRGGGQGRGVEKPPHLLAAILARGSGKGIPL 60
QY 61 KNIXHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRSSESVSKDSS 120

```

```

Db      61  KNIXHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRSSESVSKDSS 120
QY      121  TSLDAITEFLNYXEXDVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
Db      121  TSLDAITEFLNYXEXDVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
QY      181  SEIQGVRETEPLNINPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
Db      181  SEIQGVRETEPLNINPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
QY      241  ELESVDIDVDIDPIAEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
Db      241  RAERSVDIDVDIDPIAEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
QY      301  ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVIDEWRKE 360
Db      301  ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSLKLDCXEVSVSDKLAVIDEWRKE 360
QY      361  MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSQAQKAVGYICKNGRGGAIRFAEH 420
Db      361  MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSQAQKAVGYICKNGRGGAIRFAEH 420
QY      421  ICLLMKVNNSCQK 434
Db      421  ICLLMKVNNSCQK 434

RESULT 3
Q8K2G7 PRELIMINARY; PRT; 432 AA.
AC Q8K2G7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Cytidine monophospho-N-acetylneuraminic acid synthetase.
GN CMAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031500; AAH31500.1; -.
DR XCD; XCD:3337124; CMAS.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl trans.
DR Pfam; PF02348; CTP transf 3; 1.
SQ SEQUENCE 432 AA; 48030 MW; 404CEB2302AB6A47 CRC64;

Query Match 90.6%; Score 2052; DB 11; Length 432;
Best Local Similarity 91.0%; Pred. No. 2.9e-161;
Matches 395; Conservative 12; Mismatches 25; Indels 2; Gaps 1;

QY 1 MDSVEKGAATSVSNPRGSRGPPKLCNSRGGGQGRGVEKPPHLLAAILARGSGKGIPL 60
Db 1 MDALEKGAATSGAPRGSRGPPKLCNSR- -GAGRGLEKPPHLLAAILARGSGKGIPL 58
QY 61 KNIXHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRSSESVSKDSS 120
Db 59 KNIKELAGVPLIGWVLRALDAGVFQSVVWSTDHDEIENVAQFGAQVHRSSESVSKDSS 118
QY 121 TSLDAITEFLNYXEXDVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 180
Db 119 TSLDAITEFLNYXEXDVGNIQATSCXCLHPTDLOKVAEMIREEGYDSXFSVVRHQFRW 178
QY 181 SEIQGVRETEPLNINPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 240
Db 179 SEIQGVRETEPLNINPAKPRQDWDGELYENGSGFYFAKXHLIEMGYLOGGKWHHTKC 238
QY 241 ELESVDIDVDIDPIAEQRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGDQKEI 300

```

```

Db 239 RAHSHVDIDVDIDWPIAEQVRLVRFYGVGKELKEIKLVNIDGCTNGHHIYVSGDQKE- 298
QY 301 ISYDVKDAIGISLLKXSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAIVDDEWRKE 360
Db 299 ISYDVKDAIGISLLKXSGIEVRLISERACSKQTLSSKLDCKTEVSVSDKLAIVDDEWRKE 358
QY 361 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 420
Db 359 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 418
QY 421 ICLLMKVNNSCK 434
Db 419 IFLLEKVNNSCK 432

RESULT 4
Q96AX5 PRELIMINARY; PRT; 432 AA.
ID O88719 AC O88719
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CMP-N-acetylneuraminic acid synthetase (EC 2.7.7.43).
GN CMAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98356124; PubMed=9689047;
RA Mueser A.K., Eckhardt M., Potvin B., Muhlenhoff M., Stanley P.,
RA Gerardy-Schahn R.;
RT "Mammalian cytidine 5 (prime)-monophosphate N-acetylneuraminic acid
RT synthetase: A nuclear protein with evolutionarily conserved structural
RT motifs.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:9140-9145 (1998).
RL EMBL; AF062215; CAA06915.1; -.
DR MGD; MGI:1337124; Cmas.
DR GO; GO:0008781; P:N-acetylneuraminat cytidyltransferase acti...; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl trans.
DR Pfam; PF02348; CTP_transf_3; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 432 AA; 48058 MW; 3F7295535E7F6CE5 CRC64;

Query Match 90.4%; Score 2048; DB 11; Length 432;
Best Local Similarity 90.8%; Pred. No. 6.3e-161;
Matches 394; Conservative 12; Mismatches 26; Indels 2; Gaps 1;

QY 1 MDSVEKGAATSVNPRGRPSRPPKLRNSRGQGRGVEKPPHLLAALILARGSGKGIPL 60
Db 1 MDALEKAVTSGPAPRGRPSRPPKLRNSR-GAGRGLEKPPHLLAALILARGSGKGIPL 58
QY 61 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENVAQFGAQRHRSSEVSKDSS 120
Db 59 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENVAQFGAQRHRSSEVSKDSS 118
QY 121 TSLDAIIEFLNKNEXDIVGNIQATSKLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 180
Db 119 TSLDAIIEFLNKNEXDIVGNIQATSKLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 178
QY 181 SEIQGVREVTEPLNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHITTC 240
Db 179 SEIQGVREVTEPLNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHITTC 238
QY 241 ELEHSDVDIDVDIDWPIAEQVRLVRFYGVGKELKEIKLVNIDGCTNGHHIYVSGDQKEI 300
Db 239 RAHSHVDIDVDIDWPIAEQVRLVRFYGVGKELKEIKLVNIDGCTNGHHIYVSGDQKEI 298
QY 301 ISYDVKDAIGISLLKXSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLAIVDDEWRKE 360

```

```

Db 299 ISYDVKDAIGISLLKXSGIEVRLISERACSKQTLSSKLDCKTEVSVSDKLAIVDDEWRKE 358
QY 361 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 420
Db 359 MGLCWKEVAYLGNVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGGGRGAIRFAEH 418
QY 421 ICLLMKVNNSCK 434
Db 419 IFLLEKVNNSCK 432

RESULT 5
Q96AX5 PRELIMINARY; PRT; 263 AA.
ID Q96AX5 AC Q96AX5
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to CMP-N-acetylneuraminic acid synthase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RS TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016609; AAH16609.1; -.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidylyl trans.
DR Pfam; PF02348; CTP_transf_3; 1.
SQ SEQUENCE 263 AA; 29561 MW; 5A51E1F0C98AB1DB CRC64;

Query Match 57.3%; Score 1299; DB 4; Length 263;
Best Local Similarity 95.4%; Pred. No. 3e-99;
Matches 251; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVNPRGRPSRPPKLRNSRGQGRGVEKPPHLLAALILARGSGKGIPL 60
Db 1 MDSVEKGAATSVNPRGRPSRPPKLRNSRGQGRGVEKPPHLLAALILARGSGKGIPL 60
QY 61 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENVAQFGAQRHRSSEVSKDSS 120
Db 61 KNIKHLAGVPLIGWLRAALDSGAFQSVWVSTHDEIENVAQFGAQRHRSSEVSKDSS 120
QY 121 TSLDAIIEFLNKNEXDIVGNIQATSKLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 180
Db 121 TSLDAIIEFLNKNEXDIVGNIQATSKLHPTDLOKVAEMIREEGYDSXFSVVRHQRFW 180
QY 181 SEIQGVREVTEPLNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHITTC 240
Db 181 SEIQGVREVTEPLNPAKRPRQDWDGELYENGSGFYFAKRLHIEMGYLQGGKWHITTC 240
QY 241 ELEHSDVDIDVDIDWPIAEQVRLVRFYGVGKELKEIKLVNIDGCTNGHHIYVSGDQKEI 300
Db 241 RAHSHVDIDVDIDWPIAEQVRLVRFYGVGKELKEIKLVNIDGCTNGHHIYVSGDQKEI 298

RESULT 6
Q96AX5 PRELIMINARY; PRT; 432 AA.
ID Q96AX5 AC Q96AX5
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CMP-sialic acid synthetase.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Prclacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]

```



```

Db 120 TSPFLVREDVGVAGVVEDGCTAVTVPFGHFWRDGADFGAGGCHGVNDKAYRP 179
Qy 203 RRQMDGELYENGSGFY-----FAK-RHLIEMGYLOGGKWHHTKCELESHSVDDIDVDWDP 255
Db 180 RRQDRPQDLLETGAAYAAVAPGFRKHREF-----FGRTDLVRTDPARVLEIDDPHD-- 231
Qy 256 IAEQVRLVRYGFGKEK-----LKEIKLLVNCIDGCLTNGHHIYSGDQKEIISYDVVKDAI 309
Db 232 LAZARALA-PHFDTARPALPFAADIADAVVLDPDGTQDDRVLDITDGRFVSVHRGDL 290
Qy 310 GISLLKSGIEVRLISERACQKLSLLKLDCKMEVSDKLAVVDEWRKEMGLCWKEVA 369
Db 291 GIALRSLGLTMLILSTEVPVVAARAKKLPLVHGLIDRKLALQKWCCEQGIAPERVL 350
Qy 370 YLNEVSEBCLKXVSGAPADACSVAKVAGVICKNGGGRGAIREFA 418
Db 351 YVGNVDNDLPCFALVGVFVAVASAHDAVRGAARAVTTVGGGAVREIA 399

RESULT 11
Q82HY3 PRELIMINARY; PRT; 432 AA.
AC Q82HY3;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative transferase.
GN SAV3374.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=2:477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
RX EMBL; AP005034; BAC71086.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR003329; Cytidyl_t trans.
DR InterPro; IPR008230; Sugar_ptase.
DR Pfam; PF02348; CTP_transf_3; 1.
DR PIRSF; PIRSF06118; Sugar Ptase; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 432 AA; 45461 MW; 053CC8C8C17D94DB CRC64;

Query Match 16.6%; Score 376; DB 16; Length 432;
Best Local Similarity 27.4%; Pred. No. 1:5e-22;
Matches 118; Conservative 75; Mismatches 188; Indels 50; Gaps 9;

Qy 30 NSRGGQGRVKKPPLAAILIARGSGKGTFLKXKHLKLAGVPLIGWVLRALDAGFQSVW 89
Db 3 HSEAGCAASVRR---VLAVIPARGSGKGVPAKXHLPLVGGVPLVARAVRECRATLVTVV 59
Qy 90 VSTHDDE-ENVAKQFGAOWHRSSEVSKDSSTSLDAIEFLN-----YXNEXDIVGN:QA 144

```

```

Db 50 VSTDHAIARAAARAGAEVVLPAALAGCTATSEAAVLHAMDTHEALEGAAVDVVLVQC 119
Qy 145 TSXCLHTPTDLOKVAEMIREEGYDSKFSVVRHQRWSEI-----QKGVRVTEEL 194
Db 120 TSPFVREDVGVVHAIKVGADTALTVPFPHGVWRDADDPALGALGAERAAREGIDTL 179
Qy 195 -----NLNPAKRRPRQDWDGELYENGSGFYPAKRLHIEMGYLOG-----GKWH 237
Db 180 VTGTATSGGICINDKSFRRPRQDPQDLLETGAVY-----GMDATGFEARHFRTE 235
Qy 238 TKCELESHSVDDIDWDPIAEQRLV-----RYGFQKEKLEKILVNCIDGCLTNGH 290
Db 236 VRTDPARVLEIDDPHD--LAAARALAPLFDANRPG--ALPTAEDIDAVVLDPDGTQDDR 291
Qy 291 IYVSGDQKEIISYDVVKDAIGISILKSGIEVRLISERACSKTSLSSKLDCKMEVSDK 350
Db 292 VLIDSDGRFVSVHRGDLGIAALKKSLTMLILSTEONPVVAARAKKLPLVHGLIDR 351
Qy 351 LAVDEWRKEMGLCWKEVAYELGNEVSEBCLKRVGLSGAPADACSVAKVAGVICKNGG 410
Db 352 DLALKQWCCEQGIAPERVLVYGVNDVNDLPCFALVGVFVAVASAHDAVRGAARAVTTVP 411
Qy 411 RGAIREFAEHI 421
Db 412 DGAIRETASWI 422

RESULT 12
Q8KDAL PRELIMINARY; PRT; 234 AA.
AC Q8KDAL;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Acylneuraminate cytidyltransferase.
GN NEUA OR CT1153.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Hickson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Kolonay J.L., Yang F.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Parksey D.,
RA Vanatavan J., Khouli H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR EMBL; AB012876; AAM72386.1; -.
DR TIGR; CT1153; -.
DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0039103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003329; Cytidyl_t trans.
DR Pfam; PF02348; CTP_transf_3; 1.
KW Transferase; Nucleotidyltransferase; Complete proteome.
SQ SEQUENCE 234 AA; 26382 MW; 38B7179E43D9C206 CRC64;

Query Match 12.7%; Score 288.5; DB 16; Length 234;
Best Local Similarity 35.1%; Pred. No. 1:1e-15;
Matches 79; Conservative 27; Mismatches 86; Indels 33; Gaps 5;

Qy 46 AAILIARGSGKGTFLKXKHLKLAGVPLIGWVLRALDAGFQSVWYSTDHDEENVAKQFG 105
Db 4 AAILIARGSGKGTFLKXKHLKLAGVPLIGWVLRALDAGFQSVWYSTDHDEENVAKQFG 63
Qy 106 AQVHRSSEVSKDSSTSLDAIEFLN-----YXNEXDIVGN:QA 160

```

```

Db 64 ABEVDRPER--SGDKATSESAHLEAIVTAERYGABPETWFIQATSPRKFGDIDRAIEL 123
Qy 161 IREEGYKXFWRRHQFR-----WSEIQKGVRETEPLMLNPAKPRQDWDGELY 212
Db 124 FRLEGANSLSVTRADDLTIWEQRGDANSV-----NFYRNGRQDRPSOFI 172
Qy 213 ENGSPY-PAKHLEIMGYLGKK-----WHTTKCLEHSVDI 248
Db 173 ENGSIYLF-PSVLRFGNRCQKLSVLMF-WQIWEITIEVDL 217

RESULT 13
Q81QV0 PRELIMINARY; PRT; 187 AA.
AC Q81QV0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG32220-PA.
GN CG32220.
OS Drosophila melanogaster (fruit fly).
OC Neoptera; Insecta; Hexapoda; Arthropoda; Pterygota;
OC Eukaryota; Metazoa;
OC Ecdysozoa;
OC Ecdysozoa; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1];
RP MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthanarayanan P.G., Scher S.E., Li P.W., Hoskins A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Blazer V., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H., Slazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier J., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.V., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann M.,
RA Rosier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobbart N., Moritz J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiser K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2];
RP SEQUENCE FROM N.A.
RA Celnik S.E., Adams M.D., Krommiller B., War K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananthanarayanan P.G., Brandt R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

```

```

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse J., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Kunco J.,
RA Paclab J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3];
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.E., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak P., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4];
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5];
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003515; AAN11643.1; -.
DR FlyBase; FBgn0052220; CG32220.
SQ SEQUENCE 187 AA; 21754 MW; 0B65CDE59DEC893E CRC64;

Query Match 11.8%; Score 268; DB 5; Length 187;
Best Local Similarity 34.1%; Pred. No. 3.8e-14;
Matches 61; Conservative 33; Mismatches 77; Indels 8; Gaps 3;

Qy 81 DSGAFQSVWSTDEHLENNAKQGAOVHRSSEVSKDSTLDAITFLNXXNEXDIWG 140
Db 5 NSTCFHFWSTDDKRAIAENAKYGAIIHHPKFAEDDTFSLPAISFLDVHRSIHDA 64
Qy 141 NIQATSCXCLHPTDQKVAEMIRE-EGYDSXPSVVRHGFQWSEIQKGVRETEPLMLNPA 199
Db 65 LFQCTSVFLKTKYIQ---EAVRKPESHDCVFAAKRSHYLRW---KVVDGELMPAEFOLIS 117
Qy 200 KRPRQDQDGLVNGSPFYFAKRLHLEMGYLGKKWHTTKCLEHSVDIDWDPIAE 258
Db 118 ARPRQDQDGLVNGSPFYFAKRLHLEMGYLGKKWHTTKCLEHSVDIDWDPIAE 176

RESULT 14
Q8F5R2 PRELIMINARY; PRT; 228 AA.
AC Q8F5R2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable acylneuraminate cytidyltransferase (EC 2.7.7.43).
GN NEU1 OR LA1605
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011338; AAN48804.1; -.
DR GO; GO:0008781; F:N-acylneuraminate cytidyltransferase acti. ; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR003323; Cytidylyl_trans.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:54 ; Search time 18.0701 Seconds
(without alignments)
1239.929 Million cell updates/sec

Title: US-09-930-440B-4

Perfect score: 2266

Sequence: 1 MDSVEKGAATSVSNPRGRPS.....REFAEHICLLMEKVNNSCQK 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2258	99.6	434	US-09-516-143A-2	Sequence 2, Appli
2	225.5	10.0	169	US-09-516-914-7	Sequence 7, Appli
3	216	9.5	194	US-09-543-681A-5902	Sequence 5902, Ap
4	215.5	9.5	255	US-09-489-039A-13663	Sequence 13663, A
5	213.5	9.4	377	US-09-252-991A-24747	Sequence 24747, A
6	197.5	8.7	177	US-09-540-236-3597	Sequence 3597, Ap
7	169.5	7.5	221	US-09-495-406-23	Sequence 23, Appl
8	159	7.0	180	US-09-328-352-6372	Sequence 6372, Ap
9	128	5.6	467	US-08-867-611-28	Sequence 28, Appl
10	128	5.6	467	US-09-690-359-28	Sequence 28, Appl
11	128	5.6	467	PCT-US92-06965A-33	Sequence 33, Appl
12	128	5.6	541	US-08-867-611-8	Sequence 8, Appli
13	128	5.6	541	US-09-690-359-8	Sequence 8, Appli
14	128	5.6	541	PCT-US92-06965A-13	Sequence 13, Appl
15	128	5.6	594	US-08-867-611-48	Sequence 48, Appl
16	128	5.6	594	US-09-690-359-48	Sequence 48, Appl
17	128	5.6	599	US-08-867-611-18	Sequence 18, Appl
18	128	5.6	599	US-09-690-359-18	Sequence 18, Appl
19	128	5.6	599	PCT-US92-06965A-23	Sequence 23, Appl
20	122	5.4	496	US-08-867-611-10	Sequence 10, Appl
21	122	5.4	496	US-09-690-359-10	Sequence 10, Appl
22	122	5.4	496	PCT-US92-06965A-15	Sequence 15, Appl
23	121	5.3	387	US-08-867-611-12	Sequence 12, Appl
24	121	5.3	387	US-09-690-359-12	Sequence 12, Appl
25	121	5.3	387	PCT-US92-06965A-17	Sequence 17, Appl
26	120	5.3	460	US-08-912-129A-60	Sequence 60, Appl
27	120	5.3	466	US-08-912-129A-55	Sequence 55, Appl

28	120	5.3	490	2	US-08-912-129A-50	Sequence 50, Appl
29	120	5.3	618	2	US-08-912-129A-54	Sequence 54, Appl
30	119.5	5.3	342	3	US-08-867-611-37	Sequence 37, Appl
31	119.5	5.3	342	4	US-09-690-359-37	Sequence 37, Appl
32	119.5	5.3	344	3	US-08-867-611-38	Sequence 38, Appl
33	119.5	5.3	344	4	US-09-690-359-38	Sequence 38, Appl
34	119.5	5.3	352	3	US-08-867-611-39	Sequence 39, Appl
35	119.5	5.3	352	4	US-09-690-359-39	Sequence 39, Appl
36	119.5	5.3	357	3	US-08-867-611-40	Sequence 40, Appl
37	119.5	5.3	357	4	US-09-690-359-40	Sequence 40, Appl
38	119.5	5.3	362	3	US-08-867-611-41	Sequence 41, Appl
39	119.5	5.3	362	4	US-09-690-359-41	Sequence 41, Appl
40	119.5	5.3	363	3	US-08-867-611-43	Sequence 43, Appl
41	119.5	5.3	363	4	US-09-690-359-43	Sequence 43, Appl
42	119.5	5.3	364	3	US-08-867-611-44	Sequence 44, Appl
43	119.5	5.3	364	4	US-09-690-359-44	Sequence 44, Appl
44	119.5	5.3	365	3	US-08-867-611-42	Sequence 42, Appl
45	119.5	5.3	365	4	US-09-690-359-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-09-516-143A-2
; Sequence 2, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OR INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: P2505PCT
; CURRENT APPLICATION NUMBER: US/09/516.143A
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: 133
; OTHER INFORMATION: Xaa equals Tyr or His
; NAME/KEY: SITE
; LOCATION: 136
; OTHER INFORMATION: Xaa equals Gly or Val
; NAME/KEY: SITE
; LOCATION: 147
; OTHER INFORMATION: Xaa equals Ser or Pro
; NAME/KEY: SITE
; LOCATION: 169
; OTHER INFORMATION: Xaa equals Gly or Val
US-09-516-143A-2

Query Match 99.6%; Score 2258; DB 4; Length 434;

Best Local Similarity 100.0%; Pred. No. 1.6e-223;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MDSVEKGAATSVSNPRGRPPKLRNSRGQGRGVKPPHLLALILARGSGKIGPL	60
Db	1	MDSVEKGAATSVSNPRGRPPKLRNSRGQGRGVKPPHLLALILARGSGKIGPL	60
QY	61	KNIKHLAGVPLIGWLRAALDSGAFQSVVWVSTTDHDEIENVAKQFGAOVHRRSSSVSDSS	120
Db	61	KNIKHLAGVPLIGWLRAALDSGAFQSVVWVSTTDHDEIENVAKQFGAOVHRRSSSVSDSS	120
QY	121	TSLDAIEFLNYKXNDIVGNIQATSKLHFTDLQKVAEMIREEGYDSXFSVVRHQRW	180
Db	121	TSLDAIEFLNYKXNDIVGNIQATSKLHFTDLQKVAEMIREEGYDSXFSVVRHQRW	180
QY	181	SEIQKGVREVTPELNLNLPNAKPRQWDGELYNGSYFFAKRHLIEMGYLQGGKWHHTKC	240

Db 181 SEIQGVREVEIPLNPAKRFRQDWDGELYENGSPYFAKSHLIEMGYLQSGKWHITKC 240
 QY 241 BLEHSDVIDDIDWDIAGQVRYGFGKBLKEIKLVNIDGCLTNGHIYVSGDQKEI 300
 Db 242 BLEHSDVIDDIDWDIAGQVRYGFGKBLKEIKLVNIDGCLTNGHIYVSGDQKEI 300
 QY 301 ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLVAVDEWKE 360
 Db 301 ISYDVKDAIGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLVAVDEWKE 360
 QY 361 MGLCWKEVAYLQNEVSDDECKRVGLSGAPADACSYAQKAVGYICKNGRGGAIRFAFHH 420
 Db 361 MGLCWKEVAYLQNEVSDDECKRVGLSGAPADACSYAQKAVGYICKNGRGGAIRFAFHH 420
 QY 421 ICLLMKVNNSCK 434
 Db 421 ICLLMKVNNSCK 434

RESULT 2

US-09-516-914-7

; Sequence 7, Application US/09516914
 ; Patent No. 633340;
 ; GENERAL INFORMATION:
 ; APPLICANT: Breinig, Sabine
 ; APPLICANT: Fuchs, Georg
 ; TITLE OF INVENTION: Phenol-induced Proteins of Thauera aromatica
 ; FILE REFERENCE: EC1006 US NA
 ; CURRENT APPLICATION NUMBER: US/09/516,914
 ; EARLIER FILING DATE: 2000-03-01
 ; EARLIER APPLICATION NUMBER: 60/122,952
 ; PRIOR FILING DATE: 1999-03-05
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 7
 ; LENGTH: 169
 ; TYPE: PRT
 ; ORGANISM: Thauera aromatica
 US-09-516-914-7

Query Match 10.0%; Score 225.5; DB 4; Length 169;
 Best Local Similarity 31.2%; Pred. No. 4.8e-15;
 Matches 48; Conservative 39; Mismatches 64; Indels 3; Gaps 2;
 QY 270 EKLKEIKLVNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIEVRLISER-- 327
 Db 2 EQAKNIKLVLDVGDGVMTHGRIVINDEGIESRNFIDKGMGVIVLQCGVEVAIITSKKS 61
 QY 328 ACSKQTLSSKLDCKMEVSVSDKLVAVDEWKEKMGKWEVAYLQNEVSDDECKRVGLS 387
 Db 62 GAVRHRAEELKIK-RFHEGIIKKKTEPYAQMLEEMNISDAEVCYVGGDLDVLSMKRVGLA 120
 QY 388 GAPADACSYAQKAVGYICKNGRGGAIRFAFHHI 421
 Db 121 VAVGDAVDVKEVAAVYVTTARGGEGAVREVAELI 154

RESULT 3

US-09-543-681A-5902

; Sequence 5902, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 5902

; LENGTH: 194
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 US-09-543-681A-5902

Query Match 9.5%; Score 216; DB 4; Length 194;
 Best Local Similarity 29.8%; Pred. No. 5.8e-14;
 Matches 53; Conservative 42; Mismatches 67; Indels 16; Gaps 3;
 QY 264 IGYFGK---EKLKEIKLVNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIE 320
 Db 18 YGAVSQIIMQKAEKQQLICVDVGNISGLIYMGNGEELKAEFNRDVGIGICLLTSGIE 77
 QY 321 VRLISERACSKQTLSSKLDCKMEVSVS-----DKLVAVDEWKEKMGKWEVAYLQNE 373
 Db 78 VAIITGRQ-----SKLEDRAKTIGITYLYGQGHKLLAYQQLDITLNKPEQATYIGD 131
 QY 374 EVSDECKRVGLSGAPADACSYAQKAVGYICKNGRGGAIRFAFHHICLLMEKVNNS 431
 Db 132 DLIDLPMVEKVGSLVAVADAHPLLTPRADYVTHILGGRGAVRELCDLLLAQGLDEA 189

RESULT 4

US-09-489-039A-13663

; Sequence 13663, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2003-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 13663
 ; LENGTH: 255
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-13663

Query Match 9.5%; Score 215.5; DB 4; Length 255;
 Best Local Similarity 32.3%; Pred. No. 1e-13;
 Matches 54; Conservative 36; Mismatches 68; Indels 9; Gaps 2;
 QY 270 EKLKEIKLVNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKSGIEVRLISERAC 329
 Db 88 DRAAKIRILLILDVGDGLIYMGNGEELKAEFNRDVGIGICALTSGIEVAIITGRKA 147
 QY 330 SK-----QTLSSKLDCKMEVSVS---DKLVAVDEWKEKMGKWEVAYLQNEVSDDECKRV 384
 Db 148 KLVREDRCOTLGITHI-----YQGSQKLLAFRDLTDKLVHVPREEVAYIGDGLIDPVMVAV 203
 QY 385 GLSGAPADACSYAQKAVGYICKNGRGGAIRFAFHHICLLMEKVNNS 431
 Db 204 GLSVAVADAHPLLPRTNVTNTRINGRGCAVREVCDDLILLAQGLDEA 250

RESULT 5

US-09-252-991A-24747

; Sequence 24747, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; NUMBER OF SEQ ID NOS: 190

```

; PRIORITY: 0000
; PRIORITY CODE: 0000
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24747
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24747

```

```

Query Match      9.48; Score 213.5; DB.4; Length 377;
Best Local Similarity 28.08; Pred.No. 3.1e-13;
Matches 73; Conservative 42; Mismatches 115; Indels 31; Gaps 7;

175 RHOFRASEIOK-GYREVTEPLINLNP-----KPRQDQDGLYENGVSFYFAKRHLIEM 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 RHFRWRZASRPGSGHRRPPGCHRRPGDRPRQDRPRENSRRGSPEDHGR-----QQ 169

228 GYLQGG--KWHYTKCELEHSVDI-----DVIDDPIAEQVRVIRGYFQKEKKEIKLIVCN 281
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 DRTGGRRGRGRRPEHARPVARRSDWSDTHSAALPVR-----ARRIRIAFD 221
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 IDGCLTNHIIYVSGDQKEIISYVDKDAIGLSLKKSGIEVRLISERACSKQTLSLKLDC 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
222 VDGWLTGDKLYFVDSGSEFTFTNLDGHGIKMLIASGVRTAITIGRTPVVERARNLGI 281
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 K-NEVSVSDKLAVDWRKEMGLCWKEVAY-GNEVSDSEELCKVGLSGADPADCSYAQKA 400
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 QHLYQSRREDKLAVDELLGELGICGYEQVATGEDDLPLPVIRRVGLGNMVASADPPVROH 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

401 VGVICKONGRGGAIRFEFAHI 421
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 AHGWTAARGGEGGAARFCFELI 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db

```

RESULT 6
US-39-540-236-3597
; Sequence 3597, Application US/0954236
; Patent No. 6673910
; GENERAL INFORMATION.

APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709-2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840

```

; SEQ ID NO 3597
; LENGTH: 177

```

TYPE: PRT
ORGANISM: M. catarrhalis

US-09-540-236-3597

Query Match 8.7%; Score 197.5; DB 4; Length 177;
Best Local Similarity 29.4%; Pred. No. 4e-12;
Matches 50; Conservative 38; Mismatches 63; Indels 19

[illegible]

RESULT 7

US-09-495-406-23
; Sequence 23, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:

```

: APPLICANT: Gilbert, Michel
: APPLICANT: Makarchuk, Warren W.
: APPLICANT: National Research Council of Canada
: TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
: TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
: TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
: FILE REFERENCE: 019633-000110US
: CURRENT APPLICATION NUMBER: US/09/495,406
: CURRENT FILING DATE: 2000-01-31
: PRIOR APPLICATION NUMBER: US 60/118,213
: PRIOR FILING DATE: 1999-02-01
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 23
: LENGTH: 221
: TYPE: PRT
: ORGANISM: Campylobacter jejuni
: FEATURE:
: OTHER INFORMATION: CMP-sialic acid synthetase from C. jejuni OH4384
: OTHER INFORMATION: (ORF 10a of LOS biosynthesis locus)
: US-09-495-406-23

```

Query Match	7.5%;	Score 169.5;	DB 4;	Length 221;
Best Local Similarity	28.3%;	Pred. No. 4.3e-09;		
Matches	63;	Conservative	35;	Mismatches 106;
				Indels 19;
				Gaps 5

Qy 47 ALILRGSGSGIPLKNIKHLAGVPLIGWVLRALDSDGAFQSVWSTDHDIENVAKQFGA 106

Db 4 AIIIPARGSGGIKNKLVLLNNKPLIYTKAALNTKTSIKVYVSSDSDIILNVAKSONV 63

QY 107 QVHRRSSEVSKOSTSLDAIEFLNYXKXNDIVGNICATSCXCLHPTDQKVAEMIREEGY 166
 64 DILKRPISLQDNITSDKVLHALKYKCYBQVFLQPTSPLRNTHIDEAFNLKXNSNA 123

```

167 DSMFSVVRHQRWSEIQGV-----REWTEPELNLPKRP--RRQWDGSELYENGAFY 218
      :: |||           : : : : : : : : : : : :
124 NALISV-----SECNKLKAFVCEYEGLAGI CNDEFFWPRQKLPTTYSNGAIFY 175

```

QY 219 F-AKRHLIENGYLGGKRWHTTKCELEHSVDVDIDWPIAQ 359
 | | : | | | : | | |
 DQ 176 ILKIKELNPPSPLOQSKTKTFIMDE-SSSLDIDCLELAKKAAQ 217

RESULT 8

US-09-328-352-6372
; Sequence 6372, Application US/09328352

Patent No. 6562958
GENERAL INFORMATION

APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID

; TITLE OF INVENTION: BAUMANNI FOR DIAGNOS
 ; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04

```

; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6372

```

```

; LENGTH: 180
; TYPE: PRT

```

; ORGANISM: Acinetobacter baumannii
US-09-328-352-6372

Query Match

Best Local Similarity 23.4%; Pred. No. 3.7e-08;
Matches 39; Conservative 48; Mismatches 70; Indels

QY 270 EKUKKIKLVNIDGCTNGEIVVSGQKEIISYDVKDAGISLKKSGTEVKLISERAC 329
| : : : | | : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 9 EQARHQAQLVLDVGILSDGFVLTNTNGEIKSFDRDGLGMKLAQGAKVIITGRKS 68

QY 330 S--KQTLSSLKLDCKNEVSYSKDLAVDEWRKEMGLCWKEVAYLGNVSEDEECLKRVGLS 387

DB 69 NIVEKRXSDLGVDLWFPQ-GREDKGSALREACAQFNILPSDCIYNGDDWDTLSAFATAGMS 127

QY 388 GAPADACSYAQKAVGYICKCNGRGRAIRFEAHEICL-----LMEK 427

FILING DATE: 24-AUG-1990
 APPLICATION NUMBER: US 07/614,569
 FILING DATE: 07-NOV-1990
 APPLICATION NUMBER: US 07/748,561
 FILING DATE: 21-AUG-1991
 APPLICATION NUMBER: US 07/748,565
 FILING DATE: 21-AUG-1991
 APPLICATION NUMBER: US 07/748,566
 FILING DATE: 21-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: POREMBSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 4834.US.P6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-937-9556
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE DESCRIPTION: SEQ ID NO: 28:
 LENGTH: 467 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Query Match 5.6%; Score 128; DB 4; Length 467;
 Best Local Similarity 22.4%; Pred. No. 0.00028;
 Matches 93; Conservative 59; Mismatches 154; Indels 110; Gaps 21;

QY 48 LILARGSGKIPLNKIKHLAGVPLIGWLRALDAGAFQSVWVSTDDHDEIENVAKQFGAQ 107
 DB 6 IIPARYASTPLPGKPLVDINGKPMIVHVLERARESGA-ERIIIVATDHEDVARAVEAAGGE 64

QY 108 V-----HRSSEVSKDSSLSDAIIIEFLNYXNEXD-VGNIQATSCXCLHPTDLQKVAEMI 161
 DB 65 VCMTRADHQSGET-----RLAEVVEKCAFSDDTV-V-VNQGDPEFIPATIIQVADNL 116

QY 162 REGYDSXSFVRRHQFMRSEIQKGVREVT-----PLNLNPAKPRQDWDGELYENG-- 215
 DB 117 AQQVQKMTLAVPIHNAEAFNPAVKVW-LDAGYALYFSRATIP-----WDRDRFAEGLE 172

QY 216 ----SFYFAKRHLIEMGYLQ-----GKHHTTKCELEHSVDIDVDIOWPIAEORVLYRGY 266
 DB 173 TVGDNPL----RHIGIYGYRAGFTRRYVNWQPS--PLEH-----LEQRLVLYWG- 218

QY 267 FGKELKEIKLVNIDGCLTNGHIYVSG-----DQKEIISYDVKDAIG-----IS 312
 DB 219 ----EKI-----HVAVAQEVPGTGVDTPELDLPSTNSTMGHYPCTINYT 258

QY 313 LLK-----KSGIEVRLISERACSKQTLSSKLCKMEVSVSDKLVAVDWR-----KEMG 362
 DB 259 LFKVMYVGVGVEHRL--EVACNWTGRCDLDRDRSELSPLLSTTQWVLPSCFTTLE 316

QY 363 LCKEVAIYLGNEVSDCECKRVGLSGAPADACSYAOK-----AVGYICKC 407
 DB 317 ALTTGLIHLHQNIVDQVLYGVGSS-----IVSWAIKWEYVILLFLLLDARICSC 367

RESULT 11
 PCT-US92-06965A-33
 Sequence 33, Application PC/TUS9206965A
 GENERAL INFORMATION:
 APPLICANT: DEVARE, S.
 APPLICANT: DAILEY, S.
 TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ABBOTT LABORATORIES
 STREET: ONE ABBOTT PARK ROAD
 CITY: ABBOTT PARK
 STATE: ILLINOIS
 COUNTRY: U.S.

ZIF: 60065-3500
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/06965A
 FILING DATE: 19920821
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: POREMBSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 4834PC.02
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-937-9556
 INFORMATION FOR SEQ ID NO: 33:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 467 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US92-05965A-33

Query Match 5.6%; Score 128; DB 5; Length 467;
 Best Local Similarity 22.4%; Pred. No. 0.00028;
 Matches 93; Conservative 59; Mismatches 154; Indels 110; Gaps 21;

QY 48 LILARGSGKIPLNKIKHLAGVPLIGWLRALDAGAFQSVWVSTDDHDEIENVAKQFGAQ 107
 DB 6 IIPARYASTPLPGKPLVDINGKPMIVHVLERARESGA-ERIIIVATDHEDVARAVEAAGGE 64

QY 108 V-----HRSSEVSKDSSLSDAIIIEFLNYXNEXD-VGNIQATSCXCLHPTDLQKVAEMI 161
 DB 65 VCMTRADHQSGET-----RLAEVVEKCAFSDDTVIV-VNQGDPEFIPATIIQVADNL 116

QY 162 REGYDSXSFVRRHQFMRSEIQKGVREVT-----PLNLNPAKPRQDWDGELYENG-- 215
 DB 117 AQQVQKMTLAVPIHNAEAFNPAVKVW-LDAGYALYFSRATIP-----WDRDRFAEGLE 172

QY 216 ----SFYFAKRHLIEMGYLQ-----GKHHTTKCELEHSVDIDVDIOWPIAEORVLYRGY 266
 DB 173 TVGDNPL----RHIGIYGYRAGFTRRYVNWQPS--PLEH-----LEQRLVLYWG- 218

QY 267 FGKELKEIKLVNIDGCLTNGHIYVSG-----DQKEIISYDVKDAIG-----IS 312
 DB 219 ----EKI-----HVAVAQEVPGTGVDTPELDLPSTNSTMGHYPCTINYT 258

QY 313 LLK-----KSGIEVRLISERACSKQTLSSKLCKMEVSVSDKLVAVDWR-----KEMG 362
 DB 259 LFKVMYVGVGVEHRL--EVACNWTGRCDLDRDRSELSPLLSTTQWVLPSCFTTLE 316

QY 363 LCKEVAIYLGNEVSDCECKRVGLSGAPADACSYAOK-----AVGYICKC 407
 DB 317 ALTTGLIHLHQNIVDQVLYGVGSS-----IVSWAIKWEYVILLFLLLDARICSC 367

RESULT 12
 US-08-867-611-8
 Sequence 8, Application US/08867511
 Patent No. 6172189
 GENERAL INFORMATION:
 APPLICANT: DEVARE, SUSHIL G
 APPLICANT: DESAI, SURESH M
 APPLICANT: CASBY, JAMES M
 APPLICANT: DAILEY, STEPHEN H
 APPLICANT: DAWSON, GEORGE J
 APPLICANT: GUTIERREZ, ROBERT A
 APPLICANT: LESNIEWSKI, RICHARD R
 APPLICANT: STEWART, JAMES L
 APPLICANT: RUPPRECHT, KEVIN R
 TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT


```

: FILING DATE: 21-AUG-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/748,565
: FILING DATE: 21-AUG-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/748,566
: FILING DATE: 21-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: PCREMSKI, PRISCILLA E
: REGISTRATION NUMBER: 33,207
: REFERENCE/DOCKET NUMBER: 4834.US.P6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 708-937-6365
: TELEFAX: 708-937-9556
: INFORMATION FOR SEQ ID NO: 48:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 594 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Peptide
: US-08-867-611-48

Query Match      5.6%; Score 128; DB 3; Length 594;
Best Local Similarity 22.2%; Pred. No. 0.0041;
Matches 99; Conservative 61; Mismatches 141; Indels 144; Gaps 23;

QY 48 LILARGSKGIPLNKIKHLAGVPLIGVLRALDSGAFQSVVSTHDEIENVAKQFGAQ 107
DB 6 IIPARYASTRLPGKPLVDINGKPMIVHVLERARESGA-ERIIIVATHEDEVARAVEAAGE 64

QY 108 V-----HRRSEVSKSDSTSLDAIIEFLNYXNEXDIVGNIQATSKLHPTDLQKVAEMI 161
DB 65 VCNTPADHQSGTE-----RLAEVVEKAFSDDTIVV-NVQDEPMIPATIIQVADKL 116

QY 162 REEGYDSXFSVVRHQFRWSEIQKQVRYTE-----PLNLNPAKRPRRQDWDGELIYENG-- 215
DB 117 AQROVGMATLAVPIHNABEAFNPNAVKKVLDAGVLYPSRATIP---WDRDREAEGLIE 172

QY 216 ----SFYAKRHLIEMGYLOG-----GKWHITKCELEHSVDIDVDIPIAQORVLRVGY 266
DB 173 TWGDNFL---RHGIGYVRAGFIRYVRWQPS--PLEH---LEM-----LEQRLVLYWG- 218

QY 267 PKEKELKEIKLIVCNIDGLINGHIYVSG-----DQKEII--SYIVKDAIGISLLKES 317
DB 219 ---EKI-----HVAAQEVFGTGVDTPEDDLDPSTNSMDAHFLSQTKQS 258

QY 318 SIEV-RLISRA--CSKQ-----TLSSKLDCQKEVSVSDKLAIVVDEN 357
DB 259 SENLYLVAYQATVCARAQAAPTLLYRLGAVQNEITLTHPVTKYIMTCMSADLEVVTSTW 318

QY 358 RKEMGLCWKEVAYLCNEVSDSECLK-----RVCLSGAPA-----DAC 394
DB 319 VLVGGVLAALANY-----CLSTGCVVIYGRVLSGKPALIIPDREVLYREFDEMEEEC 369

QY 395 S-----YAKQAVGYI 404
DB 370 SQHLPVIEQXVMLEAFQFKQKALGLL 394

```

Search completed: May 6, 2004, 09:08:36
Job time : 19.0701 secs

GenCore version 5.1.6
Copyright (c) 1995 - 2004 Compugen Ltd.

OM Protein - protein search, using sw model

Run on: May 6, 2004, 09:07:45 ; Search time 46.9189 Seconds
(without alignments)
2567.492 Million cell updates/sec

Title: US-09-930-440B-4
Perfect score: 2266
Sequence: 1 MDSVERGATEVGNPRGRS.....REFAEHICLLMEKWNNSCQK 434

Scoring table: BLOSUM62

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2258	99.6	434	9	US-09-984-205-2
2	2258	99.6	434	9	US-09-930-440B-4
3	2188	95.6	448	9	US-09-925-301-1376
4	1173	51.8	245	15	US-10-012-697-1535
5	945	41.7	202	9	US-09-867-550-1194
6	376	16.6	432	14	US-10-156-761-10911
7	339	15.0	74	11	US-09-864-408A-3298
8	241	10.6	44	9	US-09-864-761-35760
9	225.5	10.0	169	9	US-09-870-162A-7
10	209.5	9.2	410	9	US-09-767-041-28
11	199.5	8.8	164	12	US-10-335-977-7689
12	177.5	7.8	209	12	US-10-335-977-7689
13	169.5	7.5	221	9	US-09-816-028A-37
14	169.5	7.5	221	14	US-10-303-161-37
15	169.5	7.5	221	14	US-10-303-118-37

16	169.5	7.5	221	14	US-10-303-128-37	Sequence 37, Appl
17	169.5	7.5	221	14	US-10-303-134-37	Sequence 37, Appl
18	169.5	7.5	221	14	US-10-303-162-37	Sequence 37, Appl
19	145	6.4	536	9	US-09-816-028A-19	Sequence 19, Appl
20	145	6.4	536	9	US-09-816-028A-25	Sequence 25, Appl
21	145	6.4	536	14	US-10-303-161-19	Sequence 19, Appl
22	145	6.4	536	14	US-10-303-161-25	Sequence 19, Appl
23	145	6.4	536	14	US-10-303-118-19	Sequence 25, Appl
24	145	6.4	536	14	US-10-303-118-25	Sequence 25, Appl
25	145	6.4	536	14	US-10-303-128-19	Sequence 13, Appl
26	145	6.4	536	14	US-10-303-128-25	Sequence 25, Appl
27	145	6.4	536	14	US-10-303-134-19	Sequence 19, Appl
28	145	6.4	536	14	US-10-303-134-25	Sequence 25, Appl
29	145	6.4	536	14	US-10-303-162-19	Sequence 19, Appl
30	145	6.4	536	14	US-10-303-162-25	Sequence 25, Appl
31	142.5	6.3	229	12	US-10-335-977-6838	Sequence 6838, Ap
32	120	5.3	460	8	US-08-911-824-60	Sequence 60, Appl
33	120	5.3	466	8	US-08-911-824-55	Sequence 55, Appl
34	120	5.3	490	8	US-08-911-824-50	Sequence 50, Appl
35	120	5.3	526	8	US-08-911-824-97	Sequence 97, Appl
36	120	5.3	618	8	US-08-911-824-54	Sequence 54, Appl
37	119.5	5.3	491	8	US-08-911-824-56	Sequence 56, Appl
38	119.5	5.3	599	8	US-08-911-824-108	Sequence 108, App
39	119.5	5.3	706	8	US-08-911-824-93	Sequence 93, Appl
40	119.5	5.3	736	8	US-08-911-824-91	Sequence 91, Appl
41	115	5.1	31	9	US-09-864-761-47912	Sequence 47912, A
42	110	4.9	99	12	US-10-335-977-6837	Sequence 6837, Ap
43	109.5	4.8	90	12	US-10-335-977-7688	Sequence 7688, Ap
44	109.5	4.8	398	9	US-09-896-852-54	Sequence 54, Appl
45	109.5	4.8	398	10	US-09-728-644-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-09-984-205-2
; Sequence 2, Application US/09984205
; Patent No. US20020137175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: EP0505D1
; CURRENT APPLICATION NUMBER: US/09/984,205
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 133
; OTHER INFORMATION: Xaa equals Tyr or His
; NAME/KEY: SITE
; LOCATION: 136
; OTHER INFORMATION: Xaa equals Gly or Val
; NAME/KEY: SITE
; LOCATION: 147
; OTHER INFORMATION: Xaa equals Ser or Pro
; NAME/KEY: SITE
; LOCATION: 169
; OTHER INFORMATION: Xaa equals Gly or Val
US-09-984-205-2

Query Match

99.6%; Score 2258; DB 9; Length 434;

Best Local Similarity 100.0%; Pred. No. 5.5e-220;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGPRSRGPPKLRNSRGQGRGVEKPPHAAALILARGSGKGIPL 60
DB 1 MDSVEKGAATSVSNPRGPRSRGPPKLRNSRGQGRGVEKPPHAAALILARGSGKGIPL 60

QY 61 KNIKHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120
DB 61 KNIKHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120

QY 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180
DB 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180

QY 181 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYPAKRLIEMGLQGGKWHHTKC 240
DB 181 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYPAKRLIEMGLQGGKWHHTKC 240

QY 241 ELEHSDVIDDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300
DB 241 ELEHSDVIDDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300

QY 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDDEWKE 360
DB 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDDEWKE 360

QY 361 MGLCWKEVAYLNGNEVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIRFAEH 420
DB 361 MGLCWKEVAYLNGNEVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIRFAEH 420

QY 421 ICLMEKVNNSCQK 434
DB 421 ICLMEKVNNSCQK 434

RESULT 2
US-09-930-440b-4
; Sequence 4, Application US/09930440B
; Patent No. US20020142386A1
; GENERAL INFORMATION:
; APPLICANT: Betenbaugh et al.
; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
; FILE REFERENCE: PF509P2
; CURRENT APPLICATION NUMBER: US/09/930,440B
; CURRENT FILING DATE: 2003-08-16
; PRIOR APPLICATION NUMBER: 60/227,579
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: 09/516,793
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/169,624
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: 60/122,582
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: The 'Xaa' at location 133 stands for His, or Tyr.
; NAME/KEY: misc feature
; LOCATION: (136)..(136)
; OTHER INFORMATION: The 'Xaa' at location 136 stands for Gly, or Val.
; NAME/KEY: misc feature
; LOCATION: (147)..(147)
; OTHER INFORMATION: The 'Xaa' at location 147 stands for Pro, or Ser.
; NAME/KEY: misc feature
; LOCATION: (169)..(169)
; OTHER INFORMATION: The 'Xaa' at location 169 stands for Gly, or Val.

US-09-930-440b-4
Query Match 99.6%; Score 2188; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 5.5e-220;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGPRSRGPPKLRNSRGQGRGVEKPPHAAALILARGSGKGIPL 60
DB 1 MDSVEKGAATSVSNPRGPRSRGPPKLRNSRGQGRGVEKPPHAAALILARGSGKGIPL 60

QY 61 KNIKHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120
DB 61 KNIKHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120

QY 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180
DB 121 TSLDAIIEFLNXXEXDIVGNQIATSCXCLHPTDLQKVAEMIREEGYDSFVSFVRRHQFHW 180

QY 181 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYPAKRLIEMGLQGGKWHHTKC 240
DB 181 SEIQGVRETEPLNPAKPRQDWDGELYENGSGFYPAKRLIEMGLQGGKWHHTKC 240

QY 241 ELEHSDVIDDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300
DB 241 ELEHSDVIDDWDPIAEQVRLRYGYFGKEKLEIKLLVCNIDGCLTNGHIYVSGQKEI 300

QY 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDDEWKE 360
DB 301 ISYDVKDAIGISLLKKSGLIEVRLISERACSKOTLSLKLDCRMEVSVSDKLAVDDEWKE 360

QY 361 MGLCWKEVAYLNGNEVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIRFAEH 420
DB 361 MGLCWKEVAYLNGNEVSDDECLKRVGLSGAPADACSYAQKAVGYICKNGRGRAIRFAEH 420

QY 421 ICLMEKVNNSCQK 434
DB 421 ICLMEKVNNSCQK 434

RESULT 3
US-09-925-301-1376
; Sequence 1376, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1376
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-1376
Query Match 96.6%; Score 2188; DB 9; Length 448;
Best Local Similarity 97.0%; Pred. No. 7.3e-213;
Matches 421; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 1 MDSVEKGAATSVSNPRGPRSRGPPKLRNSRGQGRGVEKPPHAAALILARGSGKGIPL 60
DB 15 MDSVEKGAATSVSNPRGPRSRGPPKLRNSRGQGRGVEKPPHAAALILARGSGKGIPL 74

QY 61 KNIKHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 120
DB 75 KNIKHLAGVPLIGWVLRALDSGAFQSVVWSTDHDEIENVAQFGAQVHRRSRSEVSKDSS 134

QY 121 TSLDAIEFLNXYNEXDIVGNIQATSKCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 180
DB 135 TSLDAIEFLNXYNEXDIVGNIQATSKCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 194
QY 181 SEIQKGVRETEPLNPAKRPRQDWDGSLYENGSGFYFAKRLHLEMGYLQGGKWHHTKC 240
DB 195 SEIQKGVRETEPLNPAKRPRQDWDGSLYENGSGFYFAKRLHLEMGYLQGGKWHHTKC 254
QY 241 ELESHVDIDVDIPIAEQRLRYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEI 300
DB 255 RAHSHVDIDVDIPIAEQRLRYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEI 314
QY 301 ISYDVDAAGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLVVDWEXE 360
DB 315 ISYDVDAAGISLLKSGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLVVDWEXE 374
QY 361 MGLCWEVAYLNGEVDEECLKRVGLSGAPADACSVAKVAGVYCKNGGGRGAIRFAEH 420
DB 375 MGLCWEVAYLNGEVDEECLKRVGLSGAPADACSVAKVAGVYCKNGGGRGAIRFAEH 434
QY 421 ICLLMKXVNSCQK 434
DB 435 ICLLMKXVNSCQK 448

RESULT 4

US-10-012-697-1535
; Sequence 1535, Application US/10012697
; Publication No. US20030215803A1
; GENERAL INFORMATION:

; APPLICANT: Escobedo, Jaime
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Scott, Beth
; APPLICANT: Drmanac, Radoje
; APPLICANT: Cirkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
; FILE REFERENCE: 2300-16252
; CURRENT APPLICATION NUMBER: US/10/012,697
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/254,648
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/275,668
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1535
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens

Query Match 51.8%; Score 1173; DB 15; Length 245;
Best Local Similarity 95.7%; Pred. No. 3.1e-110;
Matches 225; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 137 DIVGNIQATSKCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRWWSIQKGVRETEPLN 196
DB 3 DIVGNIQATSKCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRWWSIQKGVRETEPLN 62
QY 197 NPAKRPRQDWDGSLYENGSGFYFAKRLHLEMGYLQGGKWHHTKCELEHSDIVDIDWDPI 256
DB 63 NPAKRPRQDWDGSLYENGSGFYFAKRLHLEMGYLQGGKWHHTKCELEHSDIVDIDWDPI 122

QY 257 AEQRLRYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVDAAGISLLK 316
DB 123 AEQRLRYGYFGKELKEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVDAAGISLLK 182
QY 317 SGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLVVDWEXEGLCWEVAYL 371
DB 183 SGIEVRLISERACSKQTLSSKLDCKMEVSVSDKLVVDWEXEGLCWEVAYL 237

RESULT 5

US-09-867-550-1194
; Sequence 1194, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:

; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1194
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (7)
; OTHER INFORMATION: Wherein Xaa may be any one of Arg or Gly or Trp
; NAME/KEY: VARIANT
; LOCATION: (12)
; OTHER INFORMATION: Wherein Xaa may be any one of Ala or Arg or Gln or Glu or Gly or
; OTHER INFORMATION: Lys or Met or Pro or Ser or Thr or Val
; NAME/KEY: VARIANT
; LOCATION: (41)
; OTHER INFORMATION: Wherein Xaa may be any one of Ala or Pro or Ser or Thr

Query Match 41.7%; Score 945; DB 9; Length 202;
Best Local Similarity 92.2%; Pred. No. 3.2e-87;
Matches 188; Conservative 1; Mismatches 13; Indels 2; Gaps 1;

QY 1 MDSVEKGAATSVSNRGRPSRPPKLRNSRGGQGRGVEKPPHLLAAILLARGSGKIPL 60
DB 1 MDSVEKGAATSVSNRGRPSRPPKLRNSRGGQGRGVEKPPHLLAAILLARGSGKIPL 58
QY 61 KNIKHLAGVPLIGWVLRALDSCAFOSVWVSTHDEIENVAKQFGAQVHRSSVSKDSS 120
DB 59 KNIKHLAGVPLIGWVLRALDSCAFOSVWVSTHDEIENVAKQFGAQVHRSSVSKDSS 118
QY 121 TSLDAIEFLNXYNEXDIVGNIQATSKCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 180
DB 119 TSLDAIEFLNXYNEXDIVGNIQATSKCLHPTDLQKVAEMIREEGYDSXFSVVRHQFRW 178
QY 181 SEIQKGVRETEPLNPAKRPRR 204
DB 179 SEIQKGVRETEPLNPAKRPRR 202

RESULT 6

US-10-156-761-10911
; Sequence 10911, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI

```
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10911
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10911

Query Match      16.6%; Score 376; DB 14; Length 432;
Best Local Similarity 27.4%; Pred. No. 6.1e-29;
Matches 118; Conservative 75; Mismatches 188; Indels 50; Gaps 9;

Qy      30 NSRGQGRGVEKPPHLLAALILARGSGKGIPLKNIKHLAGVPLIGWLVRAALDSGAFQSVW 89
Db      3 HSEAGQAASVRR--VLAIVFARGSGKGVPAKNLLFVGVPVVARARECATRLVTDVV 59

Qy      90 VSTDHDEIENVAKQFGAQVHRSSSEVSKDSSTLSDAIEFILN-----YXNEXDIIVGNIQA 144
Db      60 VSTDHHAIAAAAREAGAEVVLRLPAAIAGDTATSEAAVLHAMDTHALHGAADVLLVQC 119

Qy      145 TSXCLHPTDLQVAEMIEESYDSVSVVRHQRWSEI-----QKGVREVIETPL 194
Db      120 TSPFIVREDVDGVVHAIVGKADTALTVAPEHGFVWRDADAPGALGAERAAVEGGDTL 179

Qy      195 -----NLNPAKPRQDWDGELYENGSPYFAKRLHBMGYLQG-----GKWH 237
Db      180 VTGATSGGYINHDKSFPRQRPQDLLETGAVY-----GKATGFEAREHRRFGTE 235

Qy      238 TKCELEHSVDIDVDIDPIAEQRLV-----RYGVFGKEKLKEIKLVCNIDGCLNGH 290
Db      236 VRTDPAKVLIEDPHD--LARAALAPLFDAVRPG--ALPTAESIDAVVLDGFGTQDDR 291

Qy      291 IYVSGDQKEI--SYDVKDAIGISLLKSGIEVRLISERACSKQTLSLLKLDCKMEVSVDK 350
Db      292 VLIQSDGREFSVHRGDLGIAALRKSLTLLIISTEQNPVVAARAKLK:PVLFHGDIRK 351

Qy      351 LAVVDEVRKEMGLCWKEVAYILGNEVSDECLKRVGLSGAPADACSQAQKAVGYICKNGG 410
Db      352 DLA-LKQWCEEGIAPEVRLVYGVNDVNDLPCFALVQKPVAVASAEVDVVGARAVTTVPGG 411

Qy      411 RGATREFAEHI 421
Db      412 DGAIRETASWI 422

RESULT 7
US-09-864-408A-3298
; Sequence 3298, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Shimkars, Richard A.
; TITLE OF INVENTION: NO. US20040009474A1el Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35760
; LENGTH: 44
```

```
; SEQ ID NO 3298
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-3298

Query Match      15.0%; Score 339; DB 11; Length 74;
Best Local Similarity 86.3%; Pred. No. 2.2e-26;
Matches 63; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy      204 RODWDGELYENGSPYFAKRLHBMGYLQGGKWHHTTKCELEHSVDIDVDIDPIAEQRLV 263
Db      1 RODWDGELYENGSPYFAKRLHBMGYLQGGKWHHTTKCELEHSVDIDVDIDPIAEQRLV 60

Qy      264 YGYFGKEKLKEIK 276
Db      61 FGFGKEKLKEIK 73

RESULT 8
US-09-864-761-35760
; Sequence 35760, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35760
; LENGTH: 44
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007671.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P38144, EVALUOR 6.13e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE729942.1, EVALUOR 2.00e-20
; US-09-864-761-35760

Query Match 10.6%; Score 241; DB 9; Length 44;
Best Local Similarity 100.0%; Pred. No. 8.7e-17;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 REVTEPLNLPKPRQDMGELYENGSGFYFAKRLHLEMGYLQ 231
Db 1 REVTEPLNLPKPRQDMGELYENGSGFYFAKRLHLEMGYLQ 44

RESULT 9
US-09-870-162A-7
; Sequence 7, Application US/09870162A
; Patent No. US20020042118A1
; GENERAL INFORMATION:
; APPLICANT: Breinig, Sabine
; APPLICANT: Fuchs, Georg
; TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica
; FILE REFERENCE: BC1006 US DIV
; CURRENT APPLICATION NUMBER: US/39/870,162A
; CURRENT FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 09/516914
; PRIOR FILING DATE: 2000-03-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Thauera aromatica
; US-09-870-162A-7

Query Match 10.0%; Score 225.5; DB 9; Length 169;
Best Local Similarity 31.2%; Pred. No. 2.7e-14;
Matches 48; Conservative 39; Mismatches 64; Indels 3; Gaps 2;

QY 270 EKUKELKLVNIDGCTNGHIYVSGQKEIISVDVKAIGISLLKSGIEVRLISER-- 327
Db 2 EQAKNIKLVLDVGVMTDGRIVINDEGIESRNFIDKGMGVIVLQCGVEVAITSKS 61
QY 328 ACSKQTLSSKLCKMEVSVSDKLAVDNEKMKGLCKEVAIVLGNVSDDECLKRVGLS 387
Db 62 GAVHRAEELKIK-RFHEGKIKKTEPVAQMLDEWNISDAEVCYVSDLDVLSMKRVGLA 120
QY 388 GAPADACSYAQKAVGYICKNGRGGAIRFAEHI 421
Db 121 VAVGDADVKEVAAYVTARGGHGAVREVAELI 154

RESULT 10
US-09-767-041-28
; Sequence 28, Application US/09767041
; Patent No. US20020055168A
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22

```

```

; PRIOR APPLICATION NUMBER: PCT/NL99/0046C
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98020465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98020467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: CPS2T
; US-09-767-041-28

Query Match 9.2%; Score 209.5; DB 9; Length 410;
Best Local Similarity 23.5%; Pred. No. 4.5e-12;
Matches 93; Conservative 71; Mismatches 149; Indels 83; Gaps 14;

QY 48 LILARGSGKIPLNKIKHLAGVPLIGWVIRALDSCAF--OSVWVSTDRDELENVAKQFG 105
Db 6 LIPARSGSKGLPNKNMLFLDGVPMIFHTIRAAIBSGCEKKNIIYVSTDSEVYKEICETTG 65
QY 106 AQVHRSSEVSKDSTSLDAIEPLNYNEXDIVGNIQATSKLHPTLOKVAEMIREEG 165
Db 66 VQVLMRPADLATDFTTSPQMEHFLQDPSDOOVFLLQVTSPLRSGKHVKEAMELYGKGQ 125
QY 166 YDSXFSVVRHQFRWSEIQGVRETEPLNLPKPRP-----RRQDWDGELYENGSGF 217
Db 126 AD-----HVSFTKVDKS-PTLFTLDENGFAKIOAGLGGSVRRQDEKLTLYPENGAI 176
QY 218 YFAKHLIEMGYLOGGKWHHTKCEL-----BHSVDIDVDIWPDAEQVRLRVGY----- 266
Db 177 YISSK---QAYLADKTYFSEKTAAYVMTKEDSIDVDHFDFTGVIGRIY-FDYQRREQ 231
QY 267 ---EGKEKLKEI-----KLVNCNIDG-----CLNGHIYVSGD 296
Db 232 NKPFYKRELKRLCEQRVHDSLVIGDSRLALLDGFNISIGTMTASTSLENGQGLFATP 291
QY 297 -QKEIISYDVVDKDAIGISLLKSGIEVRLISERACSK-----QTLSSKLCKMEV 345
Db 292 IKKVLISLVGNDLITDYPHEMIEDTIRQLMESLVSKAQEVETTLIAYTLFRDSVSNBEV 351
QY 346 SVSKLAVVDEWRKEMGLCKEVAIVLGNVSDDECL 381
Db 352 QLND---VIVQASSELGIGSVIDL---NEVVEKEAM 380

RESULT 11
US-10-335-977-7689
; Sequence 7689, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:

```

```
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7689:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; OR-GINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...164
; SEQUENCE DESCRIPTION: SEQ ID NO: 7689:
US-10-335-977-7689

Query Match      8.8%; Score 199.5; DB 12; Length 164;
Best Local Similarity 30.3%; Pred. No. 1.1e-11;
Matches 47; Conservative 43; Mismatches 50; Indels 15; Gaps 2;

QY 275 IKLLVNDIGLNGHIVYSGQKRTIISYDVDAIGISILKKSGIEVRLISERACSKQTL 334
DB 2 IKLLLDVDTLSDSLYFDENFHEIKAFNVKGLGMLWQKLKKIAITGR----- 54

QY 335 SSKLDCDCKME-----VVSDDKLAVDENKENGKLCWKVAYLGNVSDDECKRVGL 386
DB 55 TSIMVKRMESLGQVFGVGVENKNAVIERLKKDLQLSACEIACVGDYNDLGMFKACTL 114

QY 387 SGAPADACSYAKAVGYICKNGGRCAGTREFAEHI 421
DB 115 SFAPFAPHLKSKAYKVLQNSGGKGAVERAIDYL 149
```

```
RESULT 12
US-10-282-122A-61379
; Sequence 61379, Application US/10282122A
; Publication No. US20040329129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
```

```
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61379
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Legionella pneumophila
US-10-282-122A-61379

Query Match      7.8%; Score 177.5; DB 12; Length 209;
Best Local Similarity 29.8%; Pred. No. 2.8e-09;
Matches 61; Conservative 31; Mismatches 68; Indels 45; Gaps 6;

QY 45 LAALILARGSGKIPKNIKHLGVLGVLGWLRAALDGAFOVWVSTDHDEIENVAKOF 104
DB 3 ILAVIPARAGSKKELPKCKNRLLAGKELTAHTIVAAQLQSCCEIIVSTDSKGIADVAVOY 52

QY 105 GAQV-HRRSEVSKDSSTSLDAIIEFLNYXNEXDI-----VGNIQATSKLHPTDLQKVAE 159
DB 63 GASVPLWRSEDLATDTSVHTVIDLKFKQDMVFDSDVLLQPTSPRKPETISHAVE 122

QY 160 M'REEG-----YDSFESVVRQFQWSEIQKGVRETEPLINLPAKE 201
DB 123 IHKVTGKSVVSPISLKPSCRSIDSGNLVKPLFQDLEY-----CNENPIYX 173

QY 202 PRQWDGELY-----ENGsfy 218
DB 174 -----LNGSIVATAKQIENKsfy 193
```

```
RESULT 13
US-09-816-028A-37
; Sequence 37, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: CMP-sialic acid synthetase from C. jejuni OH4384
; OTHER INFORMATION: (ORF 10a of lipooligosaccharide (LOS) biosynthesis
; OTHER INFORMATION: locus)
US-09-816-028A-37
```

```
Query Match      7.5%; Score 169.5; DB 9; Length 221;
Best Local Similarity 28.3%; Pred. No. 2e-08;
```

Matches	63;	Conservative	35;	Mismatches	156;	Indels	19;	Gaps	5;
Qy	47	ALILARGSGKGIPLKNTKHLAGVELICGWILRAALDSGAFOSVWYSTDHDDEIRYAKOFGA	106						
De	4	ALIPARGSGKGIKNKLVLLNNKKLLIYVTIKALNTKISKVVVSSDSDEIRYAKSONV	63						
Qy	107	QVHRSRSEVKDSSTSLDAIIIEFLYNXNEXDIVGNIQATSCXCLHPTDLQKVAEMIREEGY	166						
De	64	DILKRPISLAQDNNTTSKVLHLHALFKYDVEVVFLOFTSPFIRNIHIDEAFNLYKNSNA	123						
Qy	167	DSXFSVVRRHQFRWSEIQKGW-----REVTEPLNLPKAKRP--RRDWDGELYENGFSFY	218						
De	124	NALTSV-----SECNDKILKAFVCEYGDLAGICNDEYFPMRPKLPKTYMSNGAIY	175						
Qy	219	F-AKRHLIEGYLGQGWHTTKCELSHSDVIDVIDWPIAEQ	259						
De	176	ILKTKETENPSPFOSKTKGFLMDE-SSSLDIDCLDE-KKAEQ	217						

RESULT 14

```

RECORD 14
Sequence 37, Application US/10303161
Publication No. US20030148459A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/10/303,161
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 37
LENGTH: 221
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: CMP-stalic acid synthetase from C. jejuni OH4384
OTHER INFORMATION: (ORF 19a of lipooligosaccharide (LOS) biosynthesis
OTHER INFORMATION: locus)
US-10-303-161-37

```

Query Match	7.5%;	Score 169.5;	DB 14;	Length 221;
Best Local Similarity	28.3%;	Pred. No. 2e-08;		
Matches	63;	Conservative	35;	Mismatches 106;
			Indels	19;
			Gaps	5;
Qy	47	ALILARGSGIPLKNIKHLAGVPLIGWLRLAALDSGRFQSVWVSTSDHDIENVAKQFGA	105	
Dc	4	ALIPARGSGIKKNKNIIVLNNKRLIYYTKAKUNTIS--SKVVVSSSDLELNTAKSONV	63	
Qy	107	QVHRRSEVSKDSTSLDAIEFLANYXNEXDIDVGNIOAT--XCLHPTDLOKVAEMIREEGY	166	
Dc	64	DILKRPISLAQNTTSDKVLHHLKPKYQEDVVFLQFTSPLRNTIHIDEAFNLYKNSNA	123	
Qy	167	DSXPSVVRRHQFRWSEIQGV-----REVTEPLNLFKEKP--RRQWDGELVENSFY	218	
Dc	124	NALISV-----SECDNKILKAFVNCVEYGLAGICNDEYFPWPKLIPKTYMSNGATY	175	
Qy	219	F--AKRHLEMGVIQGGKWHHTKCELEHSDVIDIDVDIPAIABQ	259	
Dc	176	ILIKIFEIPLNPSFLOSITKHFLEWDE--SSS-DICLEDLKKACQ	217	

RESULT 15

US-10-303-118-37

```

; Sequence 37, Application US/10303118
; Publication No. US20030157655A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,118
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURES:
; OTHER INFORMATION: CMP-sialic acid synthetase from C. jejuni C34384
; OTHER INFORMATION: (ORF 10a of lipooligosaccharide (LOS) Biosynthesis
; OTHER INFORMATION: locus)
; US-10-303-118-37

Query Match          7.5%; Score 169.5; DB 14; Length 221;
Best Local Similarity 28.3%; Pred. No. 2e-08;
Matches 63; Conservative 35; Mismatches 106; Indels 19; Gaps 5

Qy 47 ALTLARGSGKGIPLKNIKHLAGVELIGWLVLRALDGSAGFQSVVSTDHDEIENVAKQFCA 106
Db 4 AILPARGSGKGIKNKLVLLNNKKLLIYVYTIKAALNTRKISIKVWSSDSDEILLNYSQNV 63
Qy 107 QVHRSRSEVSKDSTSLDAIEFLNYXNEXDIVGNIQATSCXCLHPTDLCKVAEMIREBEGY 166
Db 64 DILKRPISLQADNNTSDKVLHALLFYKDYEDVVFLQPTSLRTNIHIDFAFLNLYKNSNA 123
Qy 167 DSXFSVVRHQFRWSEIQKV-----REVTEPLNLPKAPRP--RRQWDGDELYENGSGFY 218
Db 124 NALLISV-----SECNDKLLKAFVCNEYGDLAGICNDEYFPWPRQKLPXTYNSNGAIY 175
Qy 219 F-AKRHLIEWGYLOGGKWHCTKCELSHSDVIDVDWPIAQ 259
Db 176 ILAIKEFANPFSLOSKTKHFLADE-SSSLDIDCLEDLKKAQ 217

```

Search completed: May 6, 2004, 09:25:05
Job time : 47.9189 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:34 ; Search time 4393.16 Seconds
(without alignments)
10679.608 Million cell updates/sec

Title: US-09-930-440B-5

Perfect score: 1080

Sequence: 1 atgcgcgtggagc:ggagct.....gcaaaaaaatcaagtcttaa 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pi.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vi.*

29: em_htg_hum.*

30: em_htg_inv.*

31: em_htg_inv.*

32: em_htg_inv.*

33: em_htg_inv.*

34: em_htg_inv.*

35: em_htg_inv.*

36: em_htg_inv.*

37: em_htg_inv.*

38: em_sy.*

39: em_htg_hum.*

40: em_htg_mus.*

41: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1080	100.0	1080	6	BD270235	BD270235 Human gly
2	1080	100.0	1080	9	AF257466	AF257466 Homo sapi
3	1076.8	99.7	1181	9	BC019315	BC019315 Homo sapi
4	1076.8	99.7	1229	9	BC000008	BC000008 Homo sapi
5	1076.8	99.7	1230	6	AR339544	AR339544 Sequence
6	1075.2	99.6	1170	6	AX877224	AX877224 Sequence
7	1075.2	99.6	1170	6	BD156549	BD156549 Primer fo
8	1075.2	99.6	1170	9	AK001659	AK001659 Homo sapi
9	883.2	81.8	1247	10	BC003307	BC003307 Mus muscu
10	883.2	81.8	1916	10	BC057977	BC057977 Mus muscu
11	881.6	81.6	1174	10	AB041263	AB041263 Homo sapi
12	732.4	67.8	3026	9	AK091184	AK091184 Homo sapi
13	710.2	65.8	813	6	AX868777	AX868777 Sequence
14	710.2	65.8	813	6	BD148839	BD148839 Primer fo
15	483.8	44.8	514	6	AX873806	AX873806 Sequence
16	483.8	44.8	514	6	BD153868	BD153868 Primer fo
17	456.6	42.3	2373	2	AC097420	AC097420 Rattus no
18	418	38.7	464	6	AX408867	AX408867 Sequence
19	386.8	35.8	553	6	AX557002	AX557002 Sequence
20	338.6	31.4	2289	2	AC106079	AC106079 Rattus no
21	338.6	31.4	2292	2	AC094185	AC094185 Rattus no
22	280.6	26.0	404	6	AX201885	AX201885 Sequence
23	269.8	25.0	2044	9	AK126144	AK126144 Homo sapi
24	269.8	25.0	201300	9	AL137073	AL137073 Human DNA
25	241.6	22.4	2278	10	AL683884	AL683884 Mouse DNA
26	241.4	22.4	2441	2	AC097073	AC097073 Rattus no
27	221	20.5	1194	3	AF397531	AF397531 Drosophil
28	210.4	19.5	1999	9	AK123739	AK123739 Homo sapi
29	199.6	18.5	1975	2	AK137022	AK137022 Rattus no
30	195	18.1	1823	2	AC122081	AC122081 Rattus no
31	195	18.1	1918	2	AC112827	AC112827 Rattus no
32	195	18.1	2646	2	AC120718	AC120718 Rattus no
33	145.6	13.5	2921	1	SC093912	SC093912 Streptomy
34	144.4	13.4	1123	1	AB012850	AB012850 Chlorobiu
35	135	12.5	302300	1	AF005034	AF005034 Streptomy
36	126.4	11.7	298050	1	BX321861	BX321861 Nitrosomo
37	125	11.6	349652	3	AB055099	AB055099 Drosophil
38	119.8	11.1	1526	3	AB055101	AB055101 Drosophil
39	119.8	11.1	2290	3	AB055101	AB055101 Drosophil
40	119.8	11.1	2307	3	AB055100	AB055100 Drosophil
41	119.8	11.1	2344	3	AB055098	AB055098 Drosophil
42	119.8	11.1	6120	2	AC017132	AC017132 Drosophil
43	119.8	11.1	16160	3	AC007594	AC007594 Drosophil
44	119.8	11.1	22565	3	AB003695	AB003695 Drosophil
45	111.8	10.4	86896	1	RCU57682	US7682 Rhodobacter

ALIGNMENTS

RESULT 1
BD270235
LOCUS BD270235 1080 bp DNA linear PAT 17-JUL-2003
DEFINITION Human glycosylation enzymes.
ACCESSION BD270235
VERSION BD270235.1 GI:33080003
KEYWORDS JP 2002537796-A/2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1080)
AUTHORS Coleman, T.A. and Betenbaugh, M.J.
TITLE Human glycosylation enzymes
JOURNAL Patent: JP 2002537796-A 12-NOV-2002;

COMMENT	HUMAN GENOME SCIENCES INC,JOHNS HOPKINS UNIVERSITY
CS	Homo sapiens (human)
EN	JP 2002537796-A/2
ED	12-NOV-2002
PF	01-MAR-2000 JP 2000602748
PR	02-MAR-1999 US 60/122409
PI	TIMOTHY A COLEMAN,MICHAEL J BETENBAUGH
PC	C12N15/09,A61K31/711,A61K38/43,A61K48/00,A61P1/02,A61P3/00,PC A61P7/00,
PC	A61P7/02,A61P7/06,A61P11/00,A61P13/12,A61P15/00,A61P17/02,PC A61P19/02,
PC	A61P19/10,A61P21/04,A61P25/00,A61P25/16,A61P25/24,A61P25/28,
PC	A61P29/00,
PC	A61P31/18,A61P35/00,A61P37/00,A61P43/00,C07K16/40,C12N1/15,PC C12N1/19,
PC	C12N1/21,C12N5/10,C12N9/00,G01N33/15,G01N33/50,G01N33/53,PC C12N15/00,
PC	C12N5/00,A61K37/48
CC	Human glycosylation enzymes
FH	Key Location/Qualifiers
FT	CDS (1):..(1077).
FEATURES	Location/Qualifiers
source	1..1080
ORIGIN	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
Query Match	100.0%; Score 1080; DB 6; Length 1080;
Best Local Similarity	100.0%; Pred. No. 2.2e-256;
Matches 1080; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGCGCTGGTGGGCGCGGCAACACCCGCTCTTC 60
DB	1 ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGCGCTGGTGGGCGCGGCAACACCCGCTCTTC 60
QY	61 ATCAATGCGGAGATCGGCCAGAACCAACAGGCGGACCTGGACATGACGATGATGATC 120
DB	61 ATCAATGCGGAGATCGGCCAGAACCAACAGGCGGACCTGGACATGACGATGATGATC 120
QY	121 CGCATGCCCAAGAGTGTGGGCTGATTTGTCAGATTCAGAAAGTGGAGTGGATGATTC 180
DB	121 CGCATGCCCAAGAGTGTGGGCTGATTTGTCAGATTCAGAAAGTGGAGTGGATGATTC 180
QY	181 AGTTTAAATCGGAAGCTTGGAGAGCCATACACCTCGAGCATTCCTGGGGGAGAGG 240
DB	181 AGTTTAAATCGGAAGCTTGGAGAGCCATACACCTCGAGCATTCCTGGGGGAGAGG 240
QY	241 TACGGGAGCACAAAGACATCTGGAGTTCAGGATGACGATGACGATGACGATGACGAGG 300
DB	241 TACGGGAGCACAAAGACATCTGGAGTTCAGGATGACGATGACGATGACGATGACGAGG 300
QY	301 TACGGGAGGAGTTGGGATCTTTCATGCTCTGGATGATGATGATGATGATGATGATGAA 360
DB	301 TACGGGAGGAGTTGGGATCTTTCATGCTCTGGATGATGATGATGATGATGATGATGAA 360
QY	361 TCCCTGCATGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB	361 TCCCTGCATGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY	421 CCTATCTGGAAAGACAGCCAAAGGCTCGCCCAATGATGATGATGATGATGATGATGATGAT 480
DB	421 CCTATCTGGAAAGACAGCCAAAGGCTCGCCCAATGATGATGATGATGATGATGATGATGAT 480
QY	481 TCAATGGAACCATGAGGAGTTTATCAGATGATGATGATGATGATGATGATGATGATGATG 540
DB	481 TCAATGGAACCATGAGGAGTTTATCAGATGATGATGATGATGATGATGATGATGATGATG 540
QY	541 TCTCTGCACTGTACCGAGCCATACCGCTCCAGCTGAGGAGTCAACCTCGGGGTCATC 600
DB	541 TCTCTGCACTGTACCGAGCCATACCGCTCCAGCTGAGGAGTCAACCTCGGGGTCATC 600
QY	601 TCGGAATATCAGAAAGCTCTTCTGACATTCCTCATAGGATTCCTGGGATGAAACAGGC 660
Db	601 TCGGAATATCAGAAAGCTCTTCTGACATTCCTCATAGGATTCCTGGGATGAAACAGGC 660
QY	661 ATAGCGATATCTGTGGCGCAGTGGCTCTGGGGGCCAAGTGTGGAACTCACAATCT 720
Db	661 ATAGCGATATCTGTGGCGCAGTGGCTCTGGGGGCCAAGTGTGGAACTCACAATCT 720
QY	721 TTGGACAGACCTGGAAGGGAGTACCACCTCGGCTCGCTCGAGCTCGAGAACTGGCC 780
Db	721 TTGGACAGACCTGGAAGGGAGTACCACCTCGGCTCGCTCGAGCTCGAGAACTGGCC 780
QY	781 GAGCTGGTGGCTCAGTGGCTCTTGTGGAGCTGGCTCGGCTCGGCTCGGCTCGGCTCGG 840
Db	781 GAGCTGGTGGCTCAGTGGCTCTTGTGGAGCTGGCTCGGCTCGGCTCGGCTCGGCTCGG 840
QY	841 CTGGCTCTGGAGATGSCCTGCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCAAGTGA 900
Db	841 CTGGCTCTGGAGATGSCCTGCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCAAGTGA 900
QY	901 ATTCCGAGGAGCCATCTTCAATGAGAGCTGCTCACCTGAGCTGAGTGGTGGTGGTGA 960
Db	901 ATTCCGAGGAGCCATCTTCAATGAGAGCTGCTCACCTGAGCTGAGTGGTGGTGGTGA 960
QY	961 GCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGTCTCTGTCTCTCTCTCTCT 1020
Db	961 GCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGTCTCTGTCTCTCTCTCTCTCT 1020
QY	1021 GAGGATGACCATCATGGAAGATTTAGTGGGCAAGTCTCTGTCTCTCTCTCTCTCTCT 1080
Db	1021 GAGGATGACCATCATGGAAGATTTAGTGGGCAAGTCTCTGTCTCTCTCTCTCTCTCT 1080
RESULT 2	
AP257466	1080 bp mRNA linear PRI 11-JUN-2000
LOCUS	Homo sapiens N-acetylneuraminic acid phosphate synthase mRNA,
DEFINITION	complete cds.
ACCESSION	AP257466
VERSION	AP257466.1 GI:8453155
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 1080)
TITLE	Lawrence,S.M., Huddleston,K.A., Pitts,L.R., Nguyen,N., Lee,Y.C., Wain,W.F., Coleman,T.A. and Betenbaugh,M.J.
JOURNAL	Cloning and expression of the human N-acetylneuraminic acid phosphate synthase gene with 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid biosynthetic ability
MEDLINE	J Biol Chem. 275 (23), 17869-17877 (2000)
PUBMED	20298869
REFERENCE	1 (bases 1 to 1080)
AUTHORS	Lawrence,S.M., Huddleston,K.A., Pitts,L.R., Nguyen,N., Lee,Y.C., Wain,W.F., Coleman,T.A. and Betenbaugh,M.J.
TITLE	Direct Submission
JOURNAL	Submitted (18-APR-2000) Chemical Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA
FEATURES	Location/Qualifiers
source	1..1080
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/tissue_type="liver"
	1..1080
	/note="2-keto-3-deoxy-D-glycero-D-galacto-nononic acid biosynthetic ability"
	/codon_start=1
	/product="N-acetylneuraminic acid phosphate synthase"
	/protein_id="AAP75261.1"
	/db_xref="GI:8453155"
	/translation="NPLELELCPGRWVGQHPFCFIIEIGONEQGLDVAKRMIRMAK"
CDS	

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 9 Row: d Column: 7
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 12056472.

FEATURES

```

source
1..1181
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:4339 IMAGE:2821389"
/tissue_type="Lung, small cell carcinoma"
/clone_id="NIH_MGC_7"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
1..1181
/gene="RANS"
/notes="synonym: SAS"
/db_xref="LocusID:54187"
/db_xref="MIM:605202"
61..1140
/codon_start=1
/product="N-acetylneuraminic acid phosphate synthase"
/protein_id="AAH19315.1"
/db_xref="GI:17939512"
/db_xref="LocusID:54187"
/translation="MPLELELCPRWVGQHPFCFIIAETGQNHQGLDVAKMRIMAK
EGADCAKFSQSELEKFNKALDRPYTSKHSWGTGYGEKHLFSDHYRELQRYA
SEVTFAGSDGMDVFNFLPFRKVGSDTNNFYLEKTFKAGCPMVISGQM
SNTMKVQVQVKNLPNPFQCTSAVPLQEDNLRVISYQKLPDIPIGYSGE
TQALSAVALGALSERHITLDTKWTGSDHSASLEPGEIAELVRSVLEVALGSP
TKQLPCMACNEKLGKSVAVKIPBGTILLMDLTVKVGEPKGPEDIFNLVKK
VLVTVEEDITMEELVDMHGKIKS"
82..1095
/notes="SpSE; Region: Sia:ic acid synthase [Cell env:ope
biogenesis, outer membrane]"
/db_xref="CDD:COG2089"

```

misc_feature

ORIGIN

```

Query Match: 99.7%; Score 1076.8; DB 9; Length 1181;
Best Local Similarity 99.8%; Pred. No. 1.4e-255;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCGCTGGAGCTGGAGTGTGTCCGGGCGCTGGTGGGGGCGCAACACCGCTGCTTC 60
DB 61 ATGCGCTGGAGCTGGAGTGTGTCCGGGCGCTGGTGGGGGCGCAACACCGCTGCTTC 120
QY 61 ATCATTTGCCGAGATCGGCAGAACCAACACGAGCGGACCTGGAGCTAGCCAAAGCGCATGATC 120
DB 121 ATCATTTGCCGAGATCGGCAGAACCAACACGAGCGGACCTGGAGCTAGCCAAAGCGCATGATC 180
QY 121 GCGATGGCCAGGAGTGTGGGCTGATTTGTGCCAAGTTCAGAAAGTGTGAGTGAATTC 180
DB 181 GCGATGGCCAGGAGTGTGGGCTGATTTGTGCCAAGTTCAGAAAGTGTGAGTGAATTC 240
QY 181 AAGTTTATCGGAAGCCCTGGAGAGGCGATACACTCGAAGCATCTCCGGGGGAGAGCG 240
DB 241 AAGTTTATCGGAAGCCCTGGAGAGGCGATACACTCGAAGCATCTCCGGGGGAGAGCG 300
QY 241 TACGGGGAGCAACACGAGCATCTGGAGTTCAGCCATGACCATGACGAGGAGTGCAGAGG 300
DB 301 TACGGGGAGCAACACGAGCATCTGGAGTTCAGCCATGACCATGACGAGGAGTGCAGAGG 360
QY 301 TACGGGGAGGAGTGGAGTCTTTCTTCTACTGCTCTGGCATGGATGAGATGGCATGAA 360
DB 361 TACGGGGAGGAGTGGAGTCTTTCTTCTACTGCTCTGGCATGGATGAGATGGCATGAA 420
QY 361 TTCTGTGATGACTGAATGCTTCAATTTTCAAGTTGGATCTGGAGAGCTAATTAATTT 420
DB 421 TTCTGTGATGACTGAATGCTTCAATTTTCAAGTTGGATCTGGAGAGCTAATTAATTT 480
QY 421 CCTTATCTGGAAAGACAGCAAAAGCTGCCCAATGGTGTATCTCCAGTGGGATCGAG 480

```

```

DB 481 CTTATCTGGAAAGACAGACGCCAAAAGGTGGCCCAATGGTGAATCTCCAGTGGGATGCAG 540
QY 481 TCAATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 540
DB 541 TCAATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 600
QY 541 TTCTTGCACTGTACAGAGCGCATACCGCTCCAGCTGAGGAGCTCAACCTGGGGTCATC 600
DB 601 TTCTTGCACTGTACAGAGCGCATACCGCTCCAGCTGAGGAGCTCAACCTGGGGTCATC 660
QY 601 TCGGATATCAGAGCTCTTTCTGTCATTTCCATAGGGTATTCTGGGATGAAACAGGC 660
DB 661 TCGGATATCAGAGCTCTTTCTGTCATTTCCATAGGGTATTCTGGGATGAAACAGGC 720
QY 661 ATAGCGATATCTGTGGCCGAGTGGCTCTGGGGGCCAAGGTGTGGAAAGCTCACATACT 720
DB 721 ATAGCGATATCTGTGGCCGAGTGGCTCTGTGGGGGCCAAGGTGTGGAAAGCTCACATACT 780
QY 721 TTGGACACACCTGGAGGGGAGTGCACCTGGGCTCTGCTGAGGAGCTGGAGAGCTGGGC 780
DB 781 TTGGACACACCTGGAGGGGAGTGCACCTGGGCTCTGCTGAGGAGCTGGAGAGCTGGGC 840
QY 781 GAGCTGGTGGCTGAGTGGCTCTGTGGAGCGTGGCTGGGCTCCCAACCAAGCAGCTG 840
DB 841 GAGCTGGTGGCTGAGTGGCTCTGTGGAGCGTGGCTGGGCTCCCAACCAAGCAGCTG 900
QY 841 CTGCGCTGTGAGATGGCTCCATGAGAGCTGGGCAAGTCTGTGTGGGCGCAAGTGAAA 900
DB 901 CTGCGCTGTGAGATGGCTCCATGAGAGCTGGGCAAGTCTGTGTGGGCGCAAGTGAAA 960
QY 901 ATTCCGGAGGACCACTTCTACATGAGTGTGCTGCTGAGTGGTGGTGGGCGCAAG 960
DB 961 ATTCCGGAGGACCACTTCTACATGAGTGTGCTGCTGAGTGGTGGTGGGCGCAAG 1020
QY 961 GCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTGCTGCTGCTGCTGAA 1020
DB 1021 GCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTGCTGCTGCTGCTGAA 1080
QY 1021 GAGGATGACACCATGATGAGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1080
DB 1081 GAGGATGACACCATGATGAGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1140

RESULT 4
BC000008 Homo sapiens 1229 bp mRNA linear PRI 04-OCT-2003
LOCUS Homo sapiens N-acetylneuraminic acid synthase (sialic acid
DEFINITION synthase), mRNA (cdna clone MGC:934 IMAGE:3505324), complete cds.
ACCESSION BC000008
VERSION BC000008.2 GI:33875092
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1229)
Straussberg,R.B., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Scheet,R.B., Brownstein,M.J., Usdin,R.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raja,S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullah,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Mallek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Farey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,
Sanches,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,S.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E.,

```

TITLE Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 1229)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT On Aug 19, 2003 this sequence version replaced gi:12652538.
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: anadan@systemsbio.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 7 Row: f Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 12056472.

FEATURES

Location/Qualifiers
1..1229
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:934 IMAGE:3505324"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH MGC 21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1..1229
/gene="NANS"
/note="Synonym: SAS"
/db_xref="LocusID:54187"
/db_xref="WIM:605202"
79..1158
/codon_start=1
/product="N-acetylneuraminic acid phosphate synthase"
/protein_id="AAH00008.1"
/db_xref="GI:12652539"
/db_xref="LocusID:54187"
/translation="MPELELCPRVWGGHPCFIIARIGONHOGDLIDVAKRMIRWAK
EGGDAKFKQSELEFKERKALERYTSKHSWGTGYGKHLHESHQYRELQRYA
REVGIFFTASGDMDEAFELHQLNVPFFKVGSDTNFPLEYKIAKKGRPVLSGGWQ
SMTWMQVYQIVKPLNPFCFLQTSAYLQPEDNLRVISEYQKLFPPDIPIGYSQHE
TGATSAVALGAKVLEHITLDTKWSHDSASLEPGELEALVRSALRIVERALGSP
TKQLLPACMACNEKLGKSVVAKVIPEGTILTMDLTVKVGEPKYPEDIPNLVGGK
VLVTVEEDTITMEELVDNHGKKIKS"
100..1213
/note="SpSE; Region: Sialic acid synthase [Cell envelope
biogenesis, outer membrane]"
/db_xref="CDD:COG2089"

ORIGIN

Query Vatch 99.7%; Score 1076.8; DB 9; Length 1229;
Best Local Similarity 99.8%; Pred. No. 2..4e-255;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCGCGTGGAGCTGTGTCGCGGCGGTGGGTGGGGCGGCAACACCGCTGCTTC 60
|||||
DB 79 ATGCGCGTGGAGCTGTGTCGCGGCGGTGGGTGGGGCGGCAACACCGCTGCTTC 138
|||||

RESULT 5

AR339544

LOCUS

DEFINITION

Sequence 1035 from patent US 6569662.

AR339544

DNA

1230 bp

linear

PAT 17-AUG-2003

```
ACCESSION AR339544
VERSION AR339544.1 GI:33726401
KEYWORDS Unknown.
SOURCE Unclassified.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1230)
AUTHORS Tang, Y.T., Zhou, P. and Dumanac, R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 659662-A 1235 27-MAY-2003;
FEATURES Location/Qualifiers
source 1..1230
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.7%; Score 1076.8; DB 6; Length 1230;
Best Local Similarity 99.8%; Pred. No. 1.4e-255;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACCCCTGCTTC 60
Db 119 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACCCCTGCTTC 178
Qy 61 ATCATTCGCCAGATCGGCCCAACACCCAGCGGACCTGGACGTAGCCCAAGSCATGATC 120
Db 179 ATCATTCGCCAGATCGGCCCAACACCCAGCGGACCTGGACGTAGCCCAAGSCATGATC 238
Qy 121 CGCATGCCAAGAGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 180
Db 239 CGCATGCCAAGAGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 298
Qy 181 AAGTTTAAATCGGAAGCTTGGAGAGGCATACCTCGAAGCTTCTTGGGGGAAGACG 240
Db 299 AAGTTTAAATCGGAAGCTTGGAGAGGCATACCTCGAAGCTTCTTGGGGGAAGACG 358
Qy 241 TACGGGAGACAAACAGATCTCGAGTTCAGCCATGACCAAGTTCAGAGAGAGTGAGCTAGATTC 300
Db 359 TACGGGAGACAAACAGATCTCGAGTTCAGCCATGACCAAGTTCAGAGAGAGTGAGCTAGATTC 418
Qy 301 TACGCCAGGAGGTGGGATTTCTTCACTGCCCTTGGCATGGATGGATGGATGGATGGAT 360
Db 419 TACGCCAGGAGGTGGGATTTCTTCACTGCCCTTGGCATGGATGGATGGATGGATGGAT 478
Qy 361 TCCCTGATGAACTGAATGTTCATTTTCAAGTTCAGATTCGATTCGATTCGATTCGATTC 420
Db 479 TCCCTGATGAACTGAATGTTCATTTTCAAGTTCAGATTCGATTCGATTCGATTCGATTC 538
Qy 421 CCTTATCTGAAAAGACAGCCAAAAGGTCCGCCAATGATGATCTCCAGTGGGATGCAG 480
Db 539 CCTTATCTGAAAAGACAGCCAAAAGGTCCGCCAATGATGATCTCCAGTGGGATGCAG 598
Qy 481 TCAATGACACCATGAGAGAGTTTATCAGATCTGAGAGCCCTCAACCCCACTTCTGC 540
Db 599 TCAATGACACCATGAGAGAGTTTATCAGATCTGAGAGCCCTCAACCCCACTTCTGC 658
Qy 541 TCTTTCAGTGTACAGCGCATACCCCTCCAGCTGAGAGCTCAACCTTGGGATGCATC 600
Db 659 TCTTTCAGTGTACAGCGCATACCCCTCCAGCTGAGAGCTCAACCTTGGGATGCATC 718
Qy 601 TCGGAATATCAGAGCTCTTCTTCCGATTCCTTCCATAGGATTTCTGGCATGAACAGGC 660
Db 719 TCGGAATATCAGAGCTCTTCTTCCGATTCCTTCCATAGGATTTCTGGCATGAACAGGC 778
Qy 661 ATAGCGATATCTGTGGCGCAGTGGCTCTTGGGGCCAAAGGTGTGTGGAACGTCAATAC 720
Db 779 ATAGCGATATCTGTGGCGCAGTGGCTCTTGGGGCCAAAGGTGTGTGGAACGTCAATAC 838
Qy 721 TTGACAGACCTCGAAGGGAGTGAACCACTCGGCTCGCTGGAGCTCGGAGAACTGGCC 780
Db 839 TTGACAGACCTCGAAGGGAGTGAACCACTCGGCTCGCTGGAGCTCGGAGAACTGGCC 898
Qy 781 GAGCTGTGCGGTGAGTGTGCTTGTGGAGGCTGCTTGGGCTCCCAACCAAGCAGCTG 840
```

```

999 GAGCTGTGCGGTGAGTGTGCTTGTGGAGGCTGCTTGGGCTCCCAACCAAGCAGCTG 958
841 CTGCCCTGTGAGATGGCTGCAATGAGAACTGGGCAAGTCTGGTGGCCAAAGTAAA 900
959 CTGCCCTGTGAGATGGCTGCAATGAGAACTGGGCAAGTCTGGTGGCCAAAGTAAA 1018
901 ATTCCGGAAGCACCATTCTTAACATGACATGCTCACCGTGAAGGTGGGTGGGCCCCAAA 960
1019 ATTCCGGAAGCACCATTCTTAACATGACATGCTCACCGTGAAGGTGGGTGGGCCCCAAA 1078
961 GCCTATCTCTCTGGAAGCATCTTTAACTAGTGGGCAAGAGGTCTCTGCTCACTGTTCAA 1020
1079 GCCTATCTCTCTGGAAGCATCTTTAACTAGTGGGCAAGAGGTCTCTGCTCACTGTTCAA 1138
1021 GAGATGACACCATCATCGAAGAAATTTGATAGATATCATGCAAAAATAATCAAGTCTTAA 1080
1139 GAGATGACACCATCATCGAAGAAATTTGATAGATATCATGCAAAAATAATCAAGTCTTAA 1198

RESULT 6
AX877224 1170 bp DNA Linear PAT 17-DEC-2003
DEFINITION Sequence 12129 from Patent EP1074617.
ACCESSION AX877224
VERSION AX877224.1 GI:40031960
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Negai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 12129 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES Location/Qualifiers
1..1170
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
71..1150
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE89627.1"
/translating="GI:40031961"
ECGADCAKFKSELEFKFNKALDRPYTSKHSWGTGTEHRELEFSDHQVREIQRYA
BEVGIFPTAGMDMAVEFLHNLNPPFKVGGSDTNPPYLEKTAKKRPMVTSQMG
SMOTMKOVQIYKPLNPFCLQCTSAVPIQEDVNLNVISEYOKLPDIPIGYSGHE
TGIALSVAVALGTVLERTHTLDTKWKGDHSLASLEPGLASVRSVRLVERALGSP
TKQLLPCBMACNEKLKGSVVAKVXIPETGTLITMDMLTVKVGEPKGYPPEDIFNLVKK
VLVTVEEDDTIMEELVDNHHKKIXS"

ORIGIN
Query Match 99.6%; Score 1075.2; DB 6; Length 1170;
Best Local Similarity 99.7%; Pred. No. 3.4e-255;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACCCCTGCTTC 60
Db 71 ATGCCGTGGAGCTGGAGCTGTGTCGCCGGCGCTGGTGGCGGGCAACCCCTGCTTC 130
Qy 61 ATCATTCGCCAGATCGGCCCAAGAACCAACAGGCGACCTGGACGTAGCCCAAGSCATGATC 120
Db 131 ATCATTCGCCAGATCGGCCCAAGAACCAACAGGCGACCTGGACGTAGCCCAAGSCATGATC 190
Qy 121 CGCATGCCAAGAGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 180
Db 191 CGCATGCCAAGAGTGTGGGGCTGATGTGCCAAGTTCAGAGAGAGTGAGCTAGATTC 250
Qy 191 AAGTTTAAATCGGAAGCTTGGAGAGGCATATACCTCGAAGCAATTCCTGGGGGAAGACG 240
```


Db 251 AGGTTTAAATCGGAAAGCCTTGGACAGCCCATACCTCGAAGCATTCCTCGGGGAAGACG 310
 QY 241 TAGCGGAGCACAAACGACATCTGGAGTTTCCAGCCATGACAGGAGTACAGGAGTGGAGG 300
 Db 311 TAGCGGAGCACAAACGACATCTGGAGTTTCCAGCCATGACAGGAGTGGAGG 370
 QY 301 TAGCGGAGGAGGTTGGGATCTTCTTCACTGCTTGGGATGAGTGGGATGAGTGGG 360
 Db 371 TAGCGGAGGAGGTTGGGATCTTCTTCACTGCTTGGGATGAGTGGGATGAGTGGG 430
 QY 361 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 Db 431 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 490
 QY 421 CCTTATCTGGAAGACAGCCAAAGGTCGCGCCATGATGATGATGATGATGATGATG 480
 Db 491 CCTTATCTGGAAGACAGCCAAAGGTCGCGCCATGATGATGATGATGATGATGATG 550
 QY 481 TCAATGACACCATGAGGAGTGTATGATGATGATGATGATGATGATGATGATGATG 540
 Db 551 TCAATGACACCATGAGGAGTGTATGATGATGATGATGATGATGATGATGATGATG 610
 QY 541 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 Db 611 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 670
 QY 601 TCGGAATATCAGAGGATCTTCTTCTGATGATGATGATGATGATGATGATGATGATG 660
 Db 671 TCGGAATATCAGAGGATCTTCTTCTGATGATGATGATGATGATGATGATGATGATG 730
 QY 661 ATAGGATATCTGTGCGGAGTGGCTCTGGGAGGAGTGGTGGTGGTGGTGGTGGTGG 720
 Db 731 ATAGGATATCTGTGCGGAGTGGCTCTGGGAGGAGTGGTGGTGGTGGTGGTGGTGG 790
 QY 721 TTGGACAGACCTGGAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 780
 Db 791 TTGGACAGACCTGGAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 850
 QY 781 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840
 Db 851 GAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 910
 QY 841 CTGGCCCTGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
 Db 911 CTGGCCCTGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 970
 QY 901 ATCTCGGAGGACCATCTTCAATGAGGATGATGATGATGATGATGATGATGATGATG 960
 Db 971 ATCTCGGAGGACCATCTTCAATGAGGATGATGATGATGATGATGATGATGATGATG 1030
 QY 961 GCTATCTCTCTGAGACATCTTCAATGAGGATGATGATGATGATGATGATGATGATG 1020
 Db 1031 GCTATCTCTCTGAGACATCTTCAATGAGGATGATGATGATGATGATGATGATGATG 1090
 QY 1021 GAGGATGACACCATGAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1080
 Db 1091 GAGGATGACACCATGAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1150

RESULT 7
 BD156549 1170 bp DNA linear PAT 17-JAN-2003
 LOCUS
 DEFINITION
 Primer for synthesizing full-length cDNA and use thereof.
 BD156549
 ACCESSION
 VERSION
 BD156549.1 GI:27862307
 KEYWORDS
 JP 2002191363-A/11392.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 1170)
 AUTHORS
 Ota.T., Isogai.T., Nishikawa.T., Hayashi.K., Saito.K., Yamamoto.J.,

Isihii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002191363-A 11392 09-JUL-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002191363-A/11392
 PD 09-JUL-2002
 PF 28-JUL-2000 JP 2000280990
 PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAOJUN
 PI SAITO,
 PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
 PI KEIICHI NAGAI,TETSUO OTSUKI
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof PH Key
 Location/Qualifiers
 FT CDS (71)..(1147).
 source
 1..1170
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Query Match 99.6%; Score 1075.2; DB 6; Length 1170;
 Best Local Similarity 99.7%; Pred. No. 3.4e-255;
 Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ATCCGCTGAGCTGGAGCTGTGTCGCCGGGCGCTGGGCGGCGGCGGCGGCGGCGGCGG 60
 Db 71 ATCCGCTGAGCTGGAGCTGTGTCGCCGGGCGCTGGGCGGCGGCGGCGGCGGCGGCGG 130
 QY 61 ATCAATGCTCCAGATCGGCCAGAAACCCAGCGGACCTGGACGATGATGATGATGATGATG 120
 Db 131 ATCAATGCTCCAGATCGGCCAGAAACCCAGCGGACCTGGACGATGATGATGATGATG 190
 QY 121 CGCATGCCAAGAGTGTGGGCTGTGTTGCCAAGTTCAGAGTTCAGAGTTCAGAGTTCAG 180
 Db 191 CGCATGCCAAGAGTGTGGGCTGTGTTGCCAAGTTCAGAGTTCAGAGTTCAGAGTTCAG 250
 QY 181 AAGTTTAAATCGGAAAGCCTTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 Db 251 AAGTTTAAATCGGAAAGCCTTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 310
 QY 241 TAGCGGAGCACAAACGACATCTGGAGTTTCCAGGATGATGATGATGATGATGATGATG 300
 Db 311 TAGCGGAGCACAAACGACATCTGGAGTTTCCAGGATGATGATGATGATGATGATGATG 370
 QY 301 TAGCGGAGGAGTGGGATCTTCTTCACTGCTTGGGATGATGATGATGATGATGATGATG 360
 Db 371 TAGCGGAGGAGTGGGATCTTCTTCACTGCTTGGGATGATGATGATGATGATGATGATG 430
 QY 361 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 Db 431 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 490
 QY 421 CCTTATCTGGAAGACAGCCAAAGGTCGCGCCATGATGATGATGATGATGATGATGATG 480
 Db 491 CCTTATCTGGAAGACAGCCAAAGGTCGCGCCATGATGATGATGATGATGATGATGATG 550
 QY 481 TCAATGACACCATGAGGAGTGTATGATGATGATGATGATGATGATGATGATGATG 540
 Db 551 TCAATGACACCATGAGGAGTGTATGATGATGATGATGATGATGATGATGATGATG 610
 QY 541 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 Db 611 TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 670
 QY 601 TCGGAATATCAGAGGATCTTCTTCTGATGATGATGATGATGATGATGATGATGATG 660
 Db 671 TCGGAATATCAGAGGATCTTCTTCTGATGATGATGATGATGATGATGATGATGATG 730


```

QY 661 ATAGCGATATCTGCGCGCACTGGCTCTGGGGGCAAGGTGTGGAACGTCAATAACT 720
Db 731 ATAGCGATATCTGCGCGCACTGGCTCTGGGGGCAAGGTGTGGAACGTCAATAACT 790
QY 721 TTGGCAAGACCTGGAAGGGAGTGACCACTCGGCTCGCTGAGAGCTGGAGACTGGCC 780
Db 791 TTGGCAAGACCTGGAAGGGAGTGACCACTCGGCTCGCTGAGAGCTGGAGACTGGCC 850
QY 781 GAGCTGGTGGCTGAGTGGCTCTGTTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 840
Db 851 GAGCTGGTGGCTGAGTGGCTCTGTTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 910
QY 841 CTGCGCTGGTGGCTGAGTGGCTCTGTTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 900
Db 911 CTGCGCTGGTGGCTGAGTGGCTCTGTTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 970
QY 901 ATTCCGGAAGGACCACTTCTTACCAATGGACATGCTCACCGTGAAGTGGCTGAGCCCAA 960
Db 971 ATTCCGGAAGGACCACTTCTTACCAATGGACATGCTCACCGTGAAGTGGCTGAGCCCAA 1030
QY 961 GCCTATCTCTCTGAGAGACATCTTTTAACTTAGTGGGCAAGAGGTCTCTGCTCACTGTTGAA 1020
Db 1031 GCCTATCTCTCTGAGACATCTTTTAACTTAGTGGGCAAGAGGTCTCTGCTCACTGTTGAA 1090
QY 1021 GAGGATGACACATCTGGAAGATTTGGTGTAGATATCATGGCAGAAATCAAGTCTTAA 1080
Db 1091 GAGGATGACACATCTGGAAGATTTGGTGTAGATATCATGGCAGAAATCAAGTCTTAA 1150

RESULT 8
AK001659
LOCUS
DEFINITION Homo sapiens cDNA FLJ10797 fis, clone NT2RP4000657, weakly similar to SPORE COAT POLYSACCHARIDE BIOSYNTHESIS PROTEIN SPSE.
ACCESSION AK001659
VERSION AK001659.1 GI:7023052
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Aotsuka, S., Yoshikawa, Y., Matsunawa, E., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Sasaki, N.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 1170)
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- and 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES
    Location/Qualifiers
        1..1170
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="NT2RP4000657"
            /cell_line="NT2"
            /cell_type="eratocarcinoma"
            /clone_lib="NT2RP4"
            /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA)

```

```

induction."
71..1150
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91818.1"
/db_xref="GI:7023053"
/translation="MPLELELCPCRWVGQHPFCIIAETIGNHQDGLVAKMIRMAK
EGADCAKFKQSELEFKNRKAIDRPYTSKHSWGKTYGHEKHELFESHPQYRELQRYA
EYVGIFTASGMDENAVBFLHNLNPFYFKVSGDNTNNFYELKFKGPMWISSGMC
SMDTMQVYQIVKPLNPFQCTSAVPLQPDVNLRIYSYQKLFPIPIGYSGHE
TGIAISVAVALGTGLVLRHITLDTKTMKSGDHSASLEPGELAEIVSRVLRALGSP
TKQLLPCBMACNEKLGKSVAVKVPBETILLTMDMLTVKVGEPKGYPPEDIFNLVSKK
VLVTVEEDDTIMEELVDNHRKIKS"
CDS
Query Match 99.6%; Score 1075.2; DB 9; Length 1170;
Best Local Similarity 99.7%; Pred. No. 3 4e-255;
Matches 1077; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
ORIGIN
QY 1 ATGCCGCTGGAGCTGGAGCTGTGTCCCGCGCTCGCTGGGTGGCGGGCAACACCCGCTGCTTC 60
Db 71 ATGCCGCTGGAGCTGGAGCTGTGTCCCGCGCTCGCTGGGTGGCGGGCAACACCCGCTGCTTC 130
QY 61 ATCATTCGCGAGATCGCCAGAACCAACAGAGGGGCTTGGACGTAGCCCAAGCGCATGATC 120
Db 131 ATCATTCGCGAGATCGCCAGAACCAACAGAGGGGACCTGGACGTAGCCCAAGCGCATGATC 190
QY 121 CGCATGGCCAAAGGAGTGTGGGGCTGATTTGTGCCAAGTTCACAGAGAGTGCAGTAGAATTC 180
Db 191 CGCATGGCCAAAGGAGTGTGGGGCTGATTTGTGCCAAGTTCACAGAGAGTGCAGTAGAATTC 250
QY 181 AAGTTTATCGGAAGCCTTGGAGAGGCGATACACTCGAAGCATTCCTGGGGGAGACG 240
Db 251 AAGTTTATCGGAAGCCTTGGAGAGGCGATACACTCGAAGCATTCCTGGGGGAGACG 310
QY 241 TAGCGGGAGACCAACAGCATCTCGAGTTCAGCCATACACAGTACAGGAGCTGCGAGG 300
Db 311 TAGCGGGAGACCAACAGCATCTCGAGTTCAGCCATACACAGTACAGGAGCTGCGAGG 370
QY 301 TAGCCCGAGGAGTGGGATCTTCTTCACTGCTCTGGCATGGATGAGATGAGATGCA 360
Db 371 TAGCCCGAGGAGTGGGATCTTCTTCACTGCTCTGGCATGGATGAGATGAGATGCA 430
QY 361 TTCTTGCATGAACTGAATGTTCCATTTTCAAGTTGATCTGGAGACACTAATAATTTT 420
Db 431 TTCTTGCATGAACTGAATGTTCCATTTTCAAGTTGATCTGGAGACACTAATAATTTT 490
QY 421 CCTTATCTGGAAGAGACAGCCAAAGGTGCGCCATGTTGATCTCCAGTGGGATGCAG 480
Db 491 CCTTATCTGGAAGAGACAGCCAAAGGTGCGCCATGTTGATCTCCAGTGGGATGCAG 550
QY 481 TCAATGGACACCTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCGC 540
Db 551 TCAATGGACACCTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCGC 610
QY 541 TTCTTGAGTGTACAGCGGATACCCCTCCAGCTGAGGAGCTCAACCTGGGGTATC 600
Db 611 TTCTTGAGTGTACAGCGGATACCCCTCCAGCTGAGGAGCTCAACCTGGGGTATC 670
QY 501 TCGGAATATCAGAAAGCTCTTTCCTGATTCCTCAGATTCCTGAGGATTTCTGGGCATGAAACAGGC 660
Db 671 TCGGAATATCAGAAAGCTCTTTCCTGATTCCTCAGATTCCTGAGGATTTCTGGGCATGAAACAGGC 730
QY 561 ATAGCGATATCTGTCGCGAGTGGCTCTGCGGGGCCAAGGTGTTGGAACTCAATAACT 720
Db 731 ATAGCGATATCTGTCGCGAGTGGCTCTGCGGGGCCAAGGTGTTGGAACTCAATAACT 790
QY 721 TTGGACAGACCTGGAAGGGAGTGACACTCGGCGCTCGCTGGAGCCTGGAGAACTGGCC 780
Db 791 TTGGACAGACCTGGAAGGGAGTGACACTCGGCGCTCGCTGGAGCCTGGAGAACTGGCC 850
QY 781 GAGCTGGTGGCTGAGTGGCTCTGTTGGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 840

```

```

Db      851 GAGCTGGTGGCTGAGTGGTCTTGTGGAGCGTGCCCTGGCTCCCAACCAAGCAGCTG 910
QY      841 CTGCCCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGTGGCCAAAGTGAAA 900
Db      911 CTGCCCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGTGGCCAAAGTGAAA 970
QY      901 ATTCCGAAGGACCACTTCAATGAGATGCTCAGCTGAGCTGAGCTGAGCTGAGCTGAG 960
Db      971 ATTCCGAAGGACCACTTCAATGAGATGCTCAGCTGAGCTGAGCTGAGCTGAGCTGAG 1030
QY      961 GCCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGCTCTGTGCTACTGTGAA 1020
Db      1031 GCCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGCTCTGTGCTACTGTGAA 1090
QY      1021 GAGGATGACACCATCATGAGAGATTTGTAGATATCATGCGCAAAAATCAAGTCTTAA 1080
Db      1091 GAGGATGACACCATCATGAGAGATTTGTAGATATCATGCGCAAAAATCAAGTCTTAA 1150

RESULT 9
LOCUS   BC003307                1247 bp      mRNA      linear      ROD 03-OCT-2003
DEFINITION Mus musculus N-acetylneuraminic acid synthase [sialic acid
ACCESSION BC003307
VERSION   BC003307.1  GI:13097041
KEYWORDS  MGC.
SOURCE    Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1247)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.B., Brownstein, M.J., Udwin, T.B., Toshitoki, S.,
Carinci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Ketterman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Stevchenko, Y.,
Bouffard, G.O., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Schmutz, J., Myers, R.M.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
2 (bases 1 to 1247)
Straussberg, R.
Direct Submission
Submitted (20-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LJNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu

```

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX plate: 6 Row: 1 Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 16716466.

FEATURES

Location/Qualifiers

source

1..1247

/organism="Mus musculus"

/mol_type="mRNA"

/strain="129.C57BL/6J.FVB/N"

/db_xref="taxon:10090"

/clone="MGC:6982 IMAGE:3154822"

/tissue_type="Mammary tumor, Brcal-/fl; MMTV-Cre model. 10"

months old, gross tissue."

/clone_lib="NCI CGAP Mam3"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..1247

/gene="Nans"

/db_xref="LocusID:94181"

/db_xref="GI:13097042"

124..1203

/codon_start=1

/product="N-acetylneuraminic acid synthase (sialic acid synthase)"

/protein_id="AAH03307.1"

/db_xref="GI:13097042"

/db_xref="LocusID:94181"

/translation="MPLELEICPGRWGKHPCFLIAEIGNQHGDDIVAKRMITAK

EGADCAKFKSELEFKENKALRPPTKHSWGTKEHKEHLEFSDQKBLQSYA

QEIGFTASMDMAVEFHLNPPFKVSGDTPNPPYLEKTKAGRPWISSGMO

SDMTKQVYQIVKPLNPFLOCTSAYPLOPEDANLRVISEYOKLPDIPITGSGHE

TGAISVAVALGAKVLEHITLDTKWGSDHSASLPGELAEVLRSVRLVERALGSP

TKQLLPCEMAENKLGKSVVAKVKIPAGTTITLMDLTVKVGEKGYPPEDIFNLAKK

VLVTEEDTVMESVESHKCKKA"

145..1176

/note="SpSB; Region: Sialic acid synthase [Cell envelope

biogenesis, outer membrane]"

/db_xref="CDD:COG2089"

Query Match 81.8%; Score 883.2; DB 10; Length 1247;

Best Local Similarity 88.6%; Pred. NC. 1.2e-207;

Matches 957; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 1 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGCGCAACACCCGCTGCTTC 60

124 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGCGCAACACCCGCTGCTTC 183

QY 61 ATCATTCGCCAGATCGCCGAGAACCCACGAGGCGAAGCTGGACCTAGCCAAAGCGCATGTC 120

184 ATCATTCGCCAGATCGCCGAGAACCCACGAGGCGAAGCTGGACCTAGCCAAAGCGCATGTC 243

QY 121 CGCATGCCCAAGAGTGTGGGCGCTGATGTCGCAAGTTCGAGAGAGTGTAGCTAGATTC 180

244 CGCATGCCCAAGAGTGTGGGCGCTGATGTCGCAAGTTCGAGAGAGTGTAGCTAGATTC 303

QY 181 AAGTTTAAATCGGAAGACCTTTGGAGAGCGCCATACCTCGAAGCATTCCTGGGCGAAGACG 240

304 AAGTTTAAATCGGAAGACCTTTGGAGAGCGCCATACCTCGAAGCATTCCTGGGCGAAGACG 363

QY 241 TACGGGAGACAAACGACATCTGGAGTTCAGCCATGACGAGTACAGGAGCTSCAGAGG 300

364 TATGGGAGACAAACGAGCGGATCTTGGAAATTCAGCCAGACGAGTACAGGAGCTSCAGAGG 423

QY 301 TACGGGAGAGGTTGGGATCTTCTCACTGCTCTGGCATGGATGAGATGGCAGTTGAA 363

424 TATGGGAGAGGATCGGCATCTTCTCACTGCTCTGGCATGGATGAGATGGCAGTTGAG 483

ORIGIN

Query Match

Best Local Similarity

Matches 957; Conservative

QY 1 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGCGCAACACCCGCTGCTTC 60

124 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGCGCAACACCCGCTGCTTC 183

QY 61 ATCATTCGCCAGATCGCCGAGAACCCACGAGGCGAAGCTGGACCTAGCCAAAGCGCATGTC 120

184 ATCATTCGCCAGATCGCCGAGAACCCACGAGGCGAAGCTGGACCTAGCCAAAGCGCATGTC 243

QY 121 CGCATGCCCAAGAGTGTGGGCGCTGATGTCGCAAGTTCGAGAGAGTGTAGCTAGATTC 180

244 CGCATGCCCAAGAGTGTGGGCGCTGATGTCGCAAGTTCGAGAGAGTGTAGCTAGATTC 303

QY 181 AAGTTTAAATCGGAAGACCTTTGGAGAGCGCCATACCTCGAAGCATTCCTGGGCGAAGACG 240

304 AAGTTTAAATCGGAAGACCTTTGGAGAGCGCCATACCTCGAAGCATTCCTGGGCGAAGACG 363

QY 241 TACGGGAGACAAACGACATCTGGAGTTCAGCCATGACGAGTACAGGAGCTSCAGAGG 300

364 TATGGGAGACAAACGAGCGGATCTTGGAAATTCAGCCAGACGAGTACAGGAGCTSCAGAGG 423

QY 301 TACGGGAGAGGTTGGGATCTTCTCACTGCTCTGGCATGGATGAGATGGCAGTTGAA 363

424 TATGGGAGAGGATCGGCATCTTCTCACTGCTCTGGCATGGATGAGATGGCAGTTGAG 483

Villaion,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettunen,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dackson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywicki,M.I., Skalski,U., Smalius,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 1916)
Strausberg,R.
Direct Submission
Submitted (08-SEP-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapsb@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E.S. Consortium (LMML)
cDNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland,
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mc@nih.gov
Ahter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Harsen,N., Ho,S.L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantiroop,S., Thomas,P.J., Touchman,J.W.,
Tsurugaeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,E.,
Young,A., Zhang,L.-H. and Green,S.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/UNL at: <http://image.llnl.gov>
Series: IRAP Plate: 123 Row: f Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 16716466.

the following selection criteria: matched mRNA gi: 16/16466
 Location/Qualifiers
 1..1916

```

/strain="FVB/N"
/db xref="taxon:10090"
/clone="MGC:67938 IMAGE:5052339"
/tissue_type="Liver normal, 5 month old male mouse."
/clone_lib="NCI CGAP_Li9"
/lab host="DH10B"
/notes="vector: pCMV-SPORT6"
1

```

```

/db xref="LocusID:94151"

```

```
5. _1084
/ob_xref="MGI:2149820"
/cond_start=1
/cond_start=1
```

/product="N-acetylneuraminic acid synthase (sialic acid
 synthase)"
 /protein_id="AAHS7977.1"
 /db_xref="GI:37046732"
 /db_xref="LocusID:94181"
 /translation="MPLELELCPRWVGKHPCLIAIGNQHGQIDIVAKRMITAK
 EGACAKCAQKSELEFPKNRKLKLPFTKSHSGWTKTGEKHKLHLEFSHQYKELQSYA
 QEIGFTAGSMDMAVEFHELNVPFVKTSQGTNNPFLYEKTKAGRPVPISSGMQ
 SMGTMKWQYVQVKLPNPNFCLOCTSAYPQPDANLRLEYSOKLPFDLPFGYSGHE
 TGTIAITSAVAALCAKVLRIHITLDTKWQSGHSASLPGLSEALVRLVRLALSGSP
 TKQLLPCEMACNEKIGKASVVAKYKIPAGTTILTDMLTKVKGEPKGYPPEDIFNLAKKI
 VLVNTEDTNTVSVESKSKTKK"

misc_feature		26..1057	
Query Match		81.8%; Score 883.2; DB 10; Length 1916;	
Best Local Similarity		88.6%; Pred. No. 1.2e-207;	
Matches		957; Conservative 0; Mismatches 123; Indels 0; Gaps 0;	
ORIGIN			
QY	1	ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCCTGCTTC	60
DB	5	ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCCTGCTTC	64
QY	61	ATCATGTCGCGAGTCGCGGAGCAACACCGGGCGACCTGGAGCTGACCAAGCGCATGATC	120
DB	65	ATCATGTCGCGAGTCGCGGAGCAACACCGGGCGACCTGGAGCTGACCAAGCGCATGATC	124
QY	121	CSCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAAGTTCAGAGAGTCCAGAGAGTCCAGT	180
DB	125	CSCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAAGTTCAGAGAGTCCAGAGAGTCCAGT	184
QY	181	AAGTTTAAATCGGAAGCTTGGAGAGCCATACCTCGAAGCATTCTCGGGGGAGACG	240
DB	185	AAGTTTAAATCGGAAGCTTGGAGAGCCATATCTTCTGGAAGCATTCTCGGGGGAGACG	244
QY	241	TACGGGGAGCAACACGACATCTGGAGTTCAGCATGACCATGACAGGAGCTGCGAGAGG	300
DB	245	TATGGGAGCACAGCGGCTCTGGAATTCAGCCAGCACCAGTCAAGGAGCTGCGAGNC	304
QY	301	TACGCCGAGGAGTGTGGATCTTCTCACTCCCTCTGCGATGATGATGATGATGATGATG	360
DB	305	TATGCCGAGGATCGGCATCTTCTTCTACGTCCTCTGCGATGATGATGATGATGATGATG	364
QY	361	TTCCTGCACTGAATGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT	420
DB	365	TTCCTGCACTGAATGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT	424
QY	421	CCTATCTGGAAGAGCAGCAACAAAGGTGCGCCCAATGTTGATCTCCAGTGGGATGCGAG	480
DB	425	CCTATCTGGAAGAGCAGCAACAAAGGTGCGCCCAATGTTGATCTCCAGTGGGATGCGAG	484
QY	481	TCAATGACACCATGAGCAAGTGTATCAGATCTGTGAGCGCCCTCAACCCCACTCTGCG	540
DB	485	TCAATGACACCATGAGCAAGTGTATCAGATCTGTGAGCGCCCTCAACCCCACTCTGCG	544
QY	541	TCTTTGACGTGTACAGCGCATCCCGCTCCAGCTCAGGACGCTCAACCTCGGGGTCATC	600
DB	545	TCTTTGACGTGTACAGCGCATCCCGCTCCAGCTCAGGACGCTCAACCTCGGGGTCATC	604
QY	601	TGGAATATCAGAGCTCTTTCTGACATTCCTCAGTTCCTGAGGATTCCTGGGATGAAAG	660
DB	605	TGGAATATCAGAGCTCTTTCTGACATTCCTCAGTTCCTGAGGATTCCTGGGATGAAAG	664
QY	661	ATAGCGATATCTGTGCGCGAGTGGCTCTGGGGCGCAAGTGTGGAACCTCAATAACT	720
DB	665	ATAGCGATATCTGTGCGCGAGTGGCTCTGGGGCGCAAGTGTGGAACCTCAATAACT	724
QY	721	TTGGAAGAGCTGGAAGGGAGTGAACCTCGGCTCGCTGGAGCTGGAGAACTGGCC	780
DB	725	TTGGAAGAGCTGGAAGGGAGTGAACCTCGGCTCGCTGGAGCTGGAGAACTGGCC	784
QY	781	GAGCTGGTGGCTGAGTGGCTCTGTGGAGCGTGGCTCGGCTCGGCTCGGCTCGGCTG	840
DB	785	GAGCTGGTGGCTGAGTGGCTCTGTGGAGCGTGGCTCGGCTCGGCTCGGCTCGGCTG	844
QY	841	CTGCCCTGTGAGATGCTCTGCAATGAGAGCTGGGCAAGTCTGTGTGGCGCAAGTGAA	900
DB	845	CTGCCCTGTGAGATGCTCTGCAATGAGAGCTGGGCAAGTCTGTGTGGCGCAAGTGAA	904
QY	901	ATTCGGAGGACCACTTCTTACCAATGAGATGCTGCTCAGCGTGAAGGTGGGTGAGCCCAA	960
DB	905	ATTCGGAGGACCACTTCTTACCAATGAGATGCTGCTCAGCGTGAAGGTGGGTGAGCCCAA	964

QY	961	GCTATCTCTCTGAGAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGCTACTGTGAA	1020
DB	965	GCTATCTCTCTGAGAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGCTACTGTGAA	1024
QY	1021	GAGGATGACACCATCATGAGAGAAATGGTGTAGTATATCATGGCAAAATCAAGTCTTAA	1080
DB	1025	GAGGATGACACCATCATGAGAGAAATGGTGTAGTATATCATGGCAAAATCAAGTCTTAA	1084

AB041263	1174 bp	mRNA	linear	ROD 15-JUL-2000		
Mus musculus mRNA for N-acetylneuraminic acid 9-phosphate synthetase, complete cds.						
ACCESSION	AB041263					
VERSION	AB041263.1	GI:8978233				
KEYWORDS						
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
AUTHORS	Nakata, D., Close, B.E., Colley, K.J., Matsuda, T. and Kitajima, K.					
TITLE	Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphate synthase which does not have deaminoneuraminic acid (KDN): 9-phosphate synthase activity					
JOURNAL	Biochem. Biophys. Res. Commun. 273 (2), 642-648 (2000)					
MEDLINE	20334323					
PUBMED	10873658					
REFERENCE	2 (bases 1 to 1174)					
AUTHORS	Kitajima, K. and Nakata, D.					
TITLE	Direct Submission					
JOURNAL	Submitted (04-APR-2000) Ken Kitajima, Graduate school of bioagricultural sciences, Nagoya University, Department of applied Molecular biosciences; Chikusa-ku Huro-u-chou, Nagoya, Aichi 464-8601, Japan (E-mail: kitajima@agr.nagoya-u.ac.jp, Tel:+81-52-789-4130, Fax:+81-52-789-4128)					

FEATURES	Location/Qualifiers
source	1..1174
	/organism="Mus musculus"
	/mol_type="mRNA"
	/db_xref="taxon:10090"
	/cell_line="COWKald"
	71..1150
	/codon_start=1
	/product="N-acetylneuraminic acid 9-phosphate synthetase"
	/protein_id="BA098131.1"
	/db_xref="GI:8978234"
	/translation="MPELELCPRWVGKHPGFIIEIGNEHGGDIDVAEMIRTAKEGADCAKFOKSELEFKNRKALRPYTSKHSWGTGHEKRLHFSHSDVKELOSVAQELGTFPTSGMDMAVEFLHNLNVPFFKVGSGDTNFPVLEKTKAGRMVLSGGMQSGDITKQIVKPLNPFCLQCTSAYPQPEDANLRVSEYQKLPDIPDIPYSGHETGAISVAALGAKVLERHITLDTKWDSDHSASJEPGELAEIVSRVIRALGSPTKQLLPCWACNEKUGKSVVAKVIFAGTITLDTITVKGVEPKYEPEDIFNLGKKVLTWTEEDDTWEESESVESHSKKIKA"

CDS	
-----	--

ORIGIN	
Query Match	81.6%; Score 881.6; DB 10; Length 1174;
Best Local Similarity	88.5%; Pred. No. 2.9e-207;
Matches	956; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY	1	ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGTCGCTTC	60
DB	71	ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCGCTGGTGGGGCGGCAACACCCGTCGCTTC	130
QY	61	ATCATGTCGCGAGTCCGCCAGAGCAACACCGGGCGA CTTGGACGTAGCCAGCGCATGATC	120
DB	131	ATCATGTCGCGAGTCCGCCAGAGCAACACCGGGCGA CTTGGACGTAGCCAGCATGATC	190
QY	121	CGCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAAGTTCAGAGAGTCCAGAGAGTCCAGT	180
DB	191	CGCATGCGCAAGGAGTGTGGGGCTGATTGTGCCAAGTTCAGAGAGTCCAGAGAGTCCAGT	250

QY	181	AAGTTTAAATCGGAAGCCCTTGGAGAGGCATACACCTCGAAGCATTTCTCTGGGGGAAGACG	240
DB	251	AAGTTTAAACCGGAAGCCCTGGAGAGAGCCATATATCTTGAAGCAATCTATGGGAGAGAGC	310
QY	241	TACGGGAGGACCAACAGCACTCTGAGTTTCAGGCTGACCAAGTACAGGAGCTGCAGAGG	300
DB	311	TATGGGGAGCAACAGCGCATCTGGAAATTCAGCCAGACCCAGTACAAGGAGCTGCAGAGC	370
QY	301	TACGCCGAGGAGTTGGGAACTTTCTTCACTGCCTCTGGCATGSAATGAGATGCACCTTCAA	360
DB	371	TATGCGAGAGAGATCGGCATCTTTCTTCACTGCCTCTGSCATGSCAGATGGCAGTTGAG	430
QY	361	TTCTCTGCTGAACTGAATGTTCCATTTTCAAGTTTGAATCTGGAGACACTAATAATTTT	420
DB	431	TTTCTGCAAGCACTGAATGTTTCCCTTTTCAAGTTTGAATCTGGGAGCACTAACAATTT	490
QY	421	CCCTATCTGGAAAAGACAGCCAAAAGSTCGCCCATGTTGGTATCTCCAGTGGGATCGAG	480
DB	491	CCCTACTCTGGAAAAGACAGCCAAAGAGTCTCTTATGTTGATCTCCAGCGGGATGCGAG	550
QY	481	TCATGGAACCATGAAGCAAGTTTATCAGATCTGTGAAGCCCTTCAACCCCAACTTCTGCG	540
DB	551	TCAATGGAACCATGAAGCAAGTCTATCAGATCTGTGAAGCCCTGATATCCCACTTCTGCG	610
QY	541	TTCTTGCACTGATACAGGGCATACCCGCTTCAGCTTGAAGAGTCAAGCTCGCGGTATC	600
DB	611	TTCTCTCAATGCAACAGCGCGTACCCACTACAGCCCGAGGATGCCAACTCGCGGTCTATC	670
QY	601	TCGGAATATCAGAAGCTCTTTTCTTGACATTCCTATAGGTATTTCTGGCATGAACAGGC	660
DB	671	TCGGAATACAGAAGCTCTTTTCCGACATTTCCATCGGTATTTCCGGCAGCAGACGGGC	730
QY	661	ATAGCGATATCTGTGGCGCATGTGGCTCTGGGGCCCAAGGTGTTGGAACTCACTAACT	720
DB	731	ATGCCCATATCTGTGGCGCGCTGGCTCTGGGGCCCAAGGTGTTGGAACTCACTAACT	790
QY	721	TTGGACAAGACCTGGAAGGGAGTGAACCATCTGGCCCTCGCTGGAGCTCTGGAGACTCGCC	780
DB	791	TTGGACAAGACCTGGAAGGGAGTGAACCATCTAGCTCTGCTGGAGCTCTGGAGCTGGCA	850
QY	781	GAGCTGTGGGTCAAGTGCCTTTGTGGAGCTGCTCTGGCTCTCCCAACCAAGCAGCTG	840
DB	851	GAGCTGTGGGTCTGTGGCTCTGGTGGAGCGGCCCTGGCTCTCCCAACCAAGCAGCTG	910
QY	841	CTGCCCTGTGAGATGGCTCGAATGAGAGCTGGGCAAGTCTGTGTGGGCCCAAGTGAAA	900
DB	911	CTGCCCTGTGAGATGGCTCGAATGAGAGCTGGGCAAGTCTGTGTGGGCCCAAGTGAAA	970
QY	901	ATTCGGGAAGGCACCATTTCTAAACAATGCAATGCTCAACCTGGAAGTGGGTGAGCCCAA	960
DB	971	ATCCAGCAGCAGCACCACCTGACCTTGGACACGCTCACTGTGATGTTGGGGAGCCCAA	1030
QY	961	GCTATCTCTCTGAAGCATCTTTATCTAGTGGGCAAGAGTCTCTGTGTCTATGTTGAA	1020
DB	1031	GGGTATCTCTCTGAAGCATCTTTCAACCTAGCGGGCAAAAAGGTCTGTGTCTATCGAA	1090
QY	1021	GAGGATGACACCATCTGGAAGAAATTCGTAGATATCACTGGCAAAAATCAAGTCTTAA	1080
DB	1091	GAAGATGACACGCTCATGGAGATTCGGTGGAAAGTCAAGCAGAGAAATCAAGGCTTAA	1150
RESULT 12			
AK091184			
LOCUS	3026 bp mRNA linear PR- 15-JUL-2002		
DEFINITION	Homo sapiens cDNA FLJ33855 fis, clone CTONG2006515, highly similar to N-acetylneuraminic acid phosphate synthase.		
ACCESSION	AK091184		
VERSION	AK091184.1 GI:21749493		
KEYWORDS	c-igo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		


```

Db      71  ATCCGCTGGAGCTCGAGCTGTGTCCCGGCGCTCGTGGCGGCAACACCCGCTCTC 130
QY      61  ATCATTCGGGAGATCGGCGCAGAACACACAGGCGGACCTGGAGTASCCAGCGCATGATC 120
Db      131  ATCATTCGGGAGATCGGCGCAGAACACACAGGCGGACCTGGAGTASCCAGCGCATGATC 190
QY      121  CGCATGGCCAGGAGTGTGGGCTGATTTGTGCCAAGTTCACAGAGAGTGAAGTATTC 180
Db      191  CGCATGGCCAGGAGTGTGGGCTGATTTGTGCCAAGTTCACAGAGAGTGAAGTATTC 250
QY      181  AAGTTTATTCGGAAGCCCTTGGAGAGGCGATACACCTCGAAGCATTCCTGGGGGAGAGC 240
Db      251  AAGTTTATTCGGAAGCCCTTGGAGAGGCGATACACCTCGAAGCATTCCTGGGGGAGAGC 310
QY      241  TACGGGGAGCAGACAGCATCTGGAGTTCAGCCATGACAGGAGTGAAGGAGTGAAGG 300
Db      311  TACGGGGAGCAGACAGCATCTGGAGTTCAGCCATGACAGGAGTGAAGGAGTGAAGG 370
QY      301  TACGCCAGGAGGTTGGGATCTCTTCACTGCTCTGGCATGATGAGATGSCAGTTGAA 360
Db      371  TACGCCAGGAGGTTGGGATCTCTTCACTGCTCTGGCATGATGAGATGSCAGTTGAA 430
QY      361  TTCTGCTGATGAGTGAATGTTCCATTTTCAAGTGGATCTGGAGACACTAATATTT 420
Db      431  TTCTGCTGATGAGTGAATGTTCCATTTTCAAGTGGATCTGGAGACACTAATATTT 490
QY      421  CCTTATCTGGAAGAGCAGCAGCAAAAGGTCGCCAATGGTGTCTCCAGTGGAGTGCAG 480
Db      491  CCTTATCTGGAAGAGCAGCAGCAAAAGGTCGCCAATGGTGTCTCCAGTGGAGTGCAG 550
QY      481  TCAATGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 540
Db      551  TCAATGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 610
QY      541  TTCTGTCAGTGTACAGGCGATACCGCTCCAGCTGAGGACGTCAACCTGGGGTCAAC 600
Db      611  TTCTGTCAGTGTACAGGCGATACCGCTCCAGCTGAGGACGTCAACCTGGGGTCAAC 670
QY      601  TCGGAATATCAGAGAGCTCTTCCCTGACATTCCTCAGTGTATTCGGGCATGAACAGGC 660
Db      671  TCGGAATATCAGAGAGCTCTTCCCTGACATTCCTCAGTGTATTCGGGCATGAACAGGC 730
QY      661  ATACGATATCTGTGGCGCAGTGGC-TCTGGGGGCCAAGGTGTGGAACTGCATAAC 719
Db      731  ATACGATATCTGTGGCGCAGTGGC-TCTGGGGGCCAAGGTGTGGAACTGCATAAC-AC 789
QY      720  TTTCGACNAGACCTGGAGGGGAG 743
Db      790  TTTCGACNAGACCTGGAGGGGAG 813

```

```

RESULT 15
AX873806/c
LOCUS      AX873806          514 bp      DNA      linear      PAT 17-DEC-2003
DEFINITION Sequence 8711 from Patent: EP1074617.
ACCESSION  AX873806
VERSION     AX873806.1   GI:40028595
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1
AUTHORS     Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
            Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE       Primers for synthesising full-length cDNA and their use
JOURNAL     Patent: EP 1074617-A 8711 07-FEB-2001;
            Research Association for Biotechnology (JP)
FEATURES
            source
            .. 514
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"

```

```

ORIGIN
/db_xref="taxon:9606"
Query Match      44.8%; Score 483.8; DB 6; Length 514;
Best Local Similarity 98.2%; Pred. No. 8.6e-109;
Matches 485; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY      587  ACCTGGGGTCACTCTCGGNATATCAGAGCTCTTTCCTGACATTCCTCATAGGGTATCTCG 646
Db      514  ACNTGGGGTNNNTNGGAATATCAGAACTCTTCTTCTGACATTCCTCATAGGGTATCTCG 455
QY      647  GGCAATGAACAGGCTATAGCATATCTGTGGCCGCGAGTGGCTCTGGGGGCCAAGGTGTTGG 706
Db      454  GGCAATGAACAGGCTATAGCATATCTGTGGCCGCGAGTGGCTCTGGGGGCCAAGGTGTTGG 395
QY      707  AACGTCAATCAATCTTGGACAGACCTGGAAAGGAGTGAACCACTCGGCTCGCTGGAGC 766
Db      334  AACGTCAATCAATCTTGGACAGACCTGGAAAGGAGTGAACCACTCGGCTCGCTGGAGC 335
QY      767  CTGGAGAACTGGCGAGCTGGTGGCTCAGTCCGCTCTTGTGGAGCGTGGCTGGGCTCC 826
Db      334  CTGGAGAACTGGCGAGCTGGTGGCTCAGTCCGCTCTTGTGGAGCGTGGCTGGGCTCC 275
QY      827  CAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGCATGAGAAAGCTGGGCAAGTCTGTGG 886
Db      274  CAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGCATGAGAAAGCTGGGCAAGTCTGTGG 215
QY      887  TGGCCAAAGTGAAGATTCGGGAGGCGACATCTTAACTGAATGAATGATCCCGTGAAGG 946
Db      214  TGGCCAAAGTGAAGATTCGGGAGGCGACATCTTAACTGAATGAATGATCCCGTGAAGG 155
QY      947  TGGGTGAGCCCAAGCCCTATCTCTGAGAGACATCTTTAACTAGTGGGCAAGAGGTCC 1006
Db      154  TGGGTGAGCCCAAGCCCTATCTCTGAGAGACATCTTTAACTAGTGGGCAAGAGGTCC 95
QY      1007  TGGTCACTGTTGAAGAGGATGACACCATCATGGAAGAAATTTGGTAGATAATCATGCAAAA 1066
Db      94  TGGTCACTGTTGAAGAGGATGACACCATCATGGAAGAAATTTGGTAGATAATCATGCAAAA 35
QY      1067  AAATCAAGTCTTAA 1080
Db      34  AAATCAAGTCTTAA 21

```

Search completed: May 11, 2004, 17:15:36
Job time : 4392.16 Secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 465.643 Seconds
(without alignments)
9853.151 Million cell updates/sec

Title: US-09-930-440B-5

Perfect score: 1080

Sequence: 1 atgcgcgtggagctggagct.....gcaaaaaaatcaagtcttaa 1080

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124059041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_29Jan04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1080	100.0	1080	3	AAA37763 Human gly
2	1080	100.0	1080	4	AAB28458 Nucleotid
3	1080	100.0	1080	7	AAL53993 DNA encod
4	1080	100.0	1146	3	Aaa50569 Human sia
5	1078	99.8	1268	6	AB190630 Human pol
6	1076.8	99.7	1230	4	Aai59142 Human pol
7	1076.8	99.7	1230	8	ADB49125 Novel hum
8	1075.2	99.6	1170	4	AAH14557 Human cdn
9	998.2	83.2	1243	4	AAI60928 Human pol
10	732	67.8	1238	4	AAS33141 DNA encod
11	710.2	65.8	813	4	AAH06847 Human cdn
12	483.8	44.8	514	4	AAH11876 Human cdn
13	426.4	39.5	1417	5	AAS91837 DNA encod
14	418	38.7	464	6	ABN35016 Gene #151
15	386.8	35.8	553	6	ABK94731 Neurodege
16	371.6	34.4	486	8	ACH18008 Human adu
17	336.8	31.2	442	8	ACH1408 Human foe
18	327.8	30.4	417	8	ACH38131 Human end
19	315.8	29.2	480	8	ACH32141 Human end
20	280.6	26.0	404	5	AAH42798 Nucleotid
21	260.6	24.1	1028	5	AAS81438 DNA encod
22	256.2	23.7	410	4	AAH35557 Human col
23	198.2	18.4	1017	4	ABL29095 Drosophil

24	162.4	15.0	611	5	AAS91836 DNA encod
25	146	13.5	492	4	AAB88570 Human dig
26	119.8	11.1	3123	4	AB129094 Drosophil
27	119.8	11.1	3375	4	AB129196 Drosophil
28	112.4	10.4	26281	3	AAZ60929 Nucleotid
29	101.6	9.4	17276	7	ACA64723 Streptoco
30	100	9.3	1023	6	ABN67441 Streptoco
31	100	9.3	110000	6	ABN71527 11
32	90	8.3	3777	4	ABL29200 Drosophil
33	70.4	6.5	110000	2	AAV21209 04
34	63	5.8	1059	3	AAAS0570 Escherich
35	63	5.8	1059	4	AAH28459 Nucleotid
36	63	5.8	1059	7	AAL53994 DNA encod
37	62.8	5.8	999	4	AAAF91405 N. mening
38	62.8	5.8	1000	6	ABK37785 DNA seque
39	62.8	5.8	1050	6	ABA99765 N. mening
40	62.8	5.8	65632	3	AAAB1502 N. mening
41	62.8	5.8	110000	3	AAAB1490 00
42	62.8	5.8	349980	3	AAAF21544 N. mening
43	62	5.7	133	3	AAC98691 Human col
44	60	5.6	60	6	ABN36327 Human spl
45	59.6	5.5	984	6	ABV72354 Nucleotid

ALIGNMENTS

RESULT 1

AAA37763

ID AAA37763 standard; DNA; 1080 BP.

XX AAA37763;

XX AAA37763;

DT 04-DEC-2000 (first entry)

XX Human glycosylation enzyme clone HASAA37 coding sequence.

XX Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
immunotherapy; cosmetic surgery; metabolism; immune system disorder;
haematopoietic cell deficiency; blood coagulation disorder; asthma;
afibrinogenemia; blood platelet disorder; thrombocytopenia; necropia;
autoimmune disorder; Addison's disease; multiple sclerosis; neoplasia;
allergic encephalomyelitis; allergic reaction; organ rejection;
graft-versus-host disease; inflammation; hyperproliferative disorder;
sarcoidosis; infection; gene therapy; CMP sialic acid synthetase; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
CDS 1..1080
/*tag= a
/product= "Sialic acid synthetase"

WO200052136-A2.

PD 08-SEP-2000.

PF 01-MAR-2000; 2000WO-US005325.

PR 02-MAR-1999; 99US-01222409P.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Coleman TA;

XX WPI: 2000-572179/53.

XX P-PSDB; AAY90352.

XX New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
synthetase, sialic acid synthetase and aldolase and nucleic acids
encoding the proteins for treating e.g., immune system disorders,
microbial diseases.

Claim 4; Page 110-111; 115pp; English.

This sequence encodes a human glycosylation enzyme clone of the invention, designated HASAA37. The protein of this clone is a sialic acid synthetase. The sequences are useful as reagents for the differential identification of the tissues or cell types present in a biological sample, as immunological probes, for treating a disease or condition resulting from under expression of such polypeptide, for the detection and/or treatment of disorders involving aberrant glycolysis, e.g. cramps, myoglobinuria, and as tumour marker and/or immunotherapy targets. They may also be used to differentiate, proliferate and attract cells leading to the regeneration of tissues, to modulate mammalian characteristics (e.g. in cosmetic surgery) or mammalian metabolism affecting catabolism, anabolism processing, utilisation and energy storage, to change a mammal's mental state by influencing and as a food additive or preservative. The proteins can be used to assay protein levels in a sample, as a marker or detector of an immune system disorder, to inhibit cytokine activity, and as a vaccine. They may further be used to treat immune system or of haematopoietic cell deficiencies or disorders, blood coagulation disorders (e.g. afibrinogenemia), blood platelet disorders (e.g. thrombocytopenia), wounds resulting from trauma or surgery, autoimmune disorders (e.g. Addison's disease, multiple sclerosis, allergic encephalomyelitis), allergic reactions (e.g. asthma), organ rejection, graft-versus-host disease, inflammation, hyperproliferative disorders (e.g. neoplasia, purpura, sarcoidosis), diseases caused by viruses (e.g. hepatitis, meningitis, AIDS), bacteria and fungi (include e.g. tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia, cellulitis), and diseases caused by parasites (e.g. amoebiasis, coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis)

Sequence 1080 BP; 274 A; 261 C; 314 G; 231 T; 0 U; 0 Other;

Query Match 100.0%; Score 1080; DB 3; Length 1080;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGCGCTGGAGCTGGAGCTGTGTCCTGGGGGCTGGGGGCGGCAACCCGCTGCTC 60
 1 ATGCGCTGGAGCTGGAGCTGTGTCCTGGGGGCTGGGGGCGGCAACCCGCTGCTC 60
 61 ATCATTCGCGAGATCGGCGAGACACACAGGCGGACCTGGAGCTGATGATGATC 120
 61 ATCATTCGCGAGATCGGCGAGACACACAGGCGGACCTGGAGCTGATGATGATC 120
 121 CGCATGGCGAAGAGTGGGGCTGATGTCGCAAGTTCGAGAGAGAGTGAATTC 180
 121 CGCATGGCGAAGAGTGGGGCTGATGTCGCAAGTTCGAGAGAGAGTGAATTC 180
 181 AAGTTTAAATGGAAGCTTGGAGAGGCGCATACACCTCGAAGCATTCCTGGGGGAAG 240
 181 AAGTTTAAATGGAAGCTTGGAGAGGCGCATACACCTCGAAGCATTCCTGGGGGAAG 240
 241 TACGGGAGACAAACGACATCTGGAGTTCGAGCATGACCATGACAGGAGCTGAGAG 300
 241 TACGGGAGACAAACGACATCTGGAGTTCGAGCATGACCATGACAGGAGCTGAGAG 300
 301 TACGCCGAGAGGTGGGATCTTCTTCACTGCTCTGGCATGATGATGAGTGGAGTGA 360
 301 TACGCCGAGAGGTGGGATCTTCTTCACTGCTCTGGCATGATGATGAGTGGAGTGA 360
 361 TTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 361 TTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 421 CCTTATCTGGAAGAGACAGGCAAAAGTGGCGGCAATGGTGTCTCCAGTGGGATGCG 480
 421 CCTTATCTGGAAGAGACAGGCAAAAGTGGCGGCAATGGTGTCTCCAGTGGGATGCG 480
 481 TCAATGACACCATGAGAGAGTATATCAGATCGTGAAGCCCTCAACCCCACTTCTGCG 540
 481 TCAATGACACCATGAGAGAGTATATCAGATCGTGAAGCCCTCAACCCCACTTCTGCG 540
 541 TTCCTGAGTGTACACAGGCGATACCCGCTCCAGGCTGAGGAGCTCAACCTCGGGGTATC 600

Db 541 TTCCTGAGTGTACACAGGCGATACCCGCTCCAGCTCAGAGCTGATCACTGGGGTATC 600
 QY 601 TCGGAATATCAGAAAGCTCTTTCTGACATCCCATAGGGTATTTCTGGGCTATGAACAGGC 660
 Db 601 TCGGAATATCAGAAAGCTCTTTCTGACATCCCATAGGGTATTTCTGGGCTATGAACAGGC 660
 QY 661 ATAGGATATCTGTCGGCGAGTGGCTCTGCGGCTGAGTGGGCTGAGTGGGCTGAGTGGG 720
 Db 661 ATAGGATATCTGTCGGCGAGTGGCTCTGCGGCTGAGTGGGCTGAGTGGGCTGAGTGGG 720
 QY 721 TTGGAACAAGACTGGAAGGGAGTGCACCATCGGCTCGCTCGAGCTCGAGCTCGAGCTGGCC 780
 Db 721 TTGGAACAAGACTGGAAGGGAGTGCACCATCGGCTCGCTCGAGCTCGAGCTCGAGCTGGCC 780
 QY 781 GAGCTGGTGGCTGAGTGGCTCTGCGGCTGAGTGGGCTGAGTGGGCTGAGTGGGCTGAGTGG 840
 Db 781 GAGCTGGTGGCTGAGTGGCTCTGCGGCTGAGTGGGCTGAGTGGGCTGAGTGGGCTGAGTGG 840
 QY 841 CTGCGCTGAGATGGCTGCAATGAGAGTGGGCTGAGTGGGCTGAGTGGGCTGAGTGGGCTGAG 900
 Db 841 CTGCGCTGAGATGGCTGCAATGAGAGTGGGCTGAGTGGGCTGAGTGGGCTGAGTGGGCTGAG 900
 QY 901 ATTCCGGAAGGACCATTTCTAATGAGATGATGATGATGATGATGATGATGATGATGATG 960
 Db 901 ATTCCGGAAGGACCATTTCTAATGAGATGATGATGATGATGATGATGATGATGATGATG 960
 QY 961 GCTATTCCTCTGAGAGATCTTTAATCTAGTGGGCTGAGTGGGCTGAGTGGGCTGAGTGGG 1020
 Db 961 GCTATTCCTCTGAGAGATCTTTAATCTAGTGGGCTGAGTGGGCTGAGTGGGCTGAGTGGG 1020
 QY 1021 GAGATGACACATCTGGAAGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 Db 1021 GAGATGACACATCTGGAAGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 RESULT 2
 AAH28458 standard; DNA; 1080 BP.
 AC AAH28458;
 XX 17-SEP-2001 (first entry)
 XX Nucleotide sequence of a human sialic acid synthetase.
 DE Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
 KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
 KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
 KW vaccine; ss.
 XX Homo sapiens.
 QS Key Location/Qualifiers
 FH 1.1080
 FT CDS /*tag= a
 FT /product= "sialic acid synthetase"
 XX WO200142492-A1.
 XX 14-JUN-2001.
 XX 07-DEC-2000; 2000WO-US033136.
 XX 09-DEC-1999; 99US-0169839P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UYWY-) UNIV TEMPLE.
 XX (UYWY-) UNIV WYOMING.
 XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;
 PI Jarvis D;

XX WPI; 2001-441575/47.
 DR P-PSDB; AAB84683.
 XX
 PT Cells producing cytidine monophosphate-sialic acid and sialylated
 PT glycoprotein above endogenous levels for production of vaccines and
 PT therapeutics.
 XX
 XX Example 5; Fig 31; 182pp; English.
 XX
 CC The specification describes a method for manipulating carbohydrate
 CC processing pathways in cells of interest. The methods are used to
 CC manipulate multiple pathways involved with the sialylation reaction by
 CC using recombinant DNA technology and substrate feeding approaches to
 CC enable the production of sialylated glycoproteins in the cells. The
 CC sialylation process involves the post-translational addition of the donor
 CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific
 CC acceptor carbohydrate. The cells express at least one enzyme, selected
 CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,
 CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The
 CC cells are useful for producing complex sialylated glycoproteins in cells
 CC of interest, especially insect cells. Glycoproteins containing sialylated
 CC oligosaccharides are useful as vaccines, therapeutics and diagnostic
 CC tools. Cells producing complex sialylated glycoproteins are useful for
 CC enhancing the value of heterologous expression systems and increasing the
 CC application of heterologous cell expression products as vaccines,
 CC therapeutics and diagnostic tools as well as increasing the variety of
 CC heterologous proteins that can be produced and lowering biotechnology
 CC production costs. The present sequence encodes a human sialic acid
 CC synthetase, which is used in the method of the invention
 XX
 SQ Sequence 1080 BP; 274 A; 261 C; 314 G; 231 T; 0 U; 0 Other;

Query Match 100.0%; Score 1080; DB 4; Length 1080;
 Best Local Similarity 100.0%; Pred. NC. C;
 Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

541 TTCTTCAGTGTACACGCGCATACCGCTCCAGCTTCAGAGCTCAACTGGGGTCATC 600
 541 TTCTTCAGTGTACACGCGCATACCGCTCCAGCTTCAGAGCTCAACTGGGGTCATC 600
 601 TCGGATATCAGAGCTCTTCTTCAGCATTCCTCCATAGGTTATCTGGGATGAAGAGC 660
 601 TCGGATATCAGAGCTCTTCTTCAGCATTCCTCCATAGGTTATCTGGGATGAAGAGC 660
 661 ATAGCGATATCTGTGGCGCGCAGTGGCTCTGGGGGCAAGGCTTGGAACTCACTAACT 720
 661 ATAGCGATATCTGTGGCGCGCAGTGGCTCTGGGGGCAAGGCTTGGAACTCACTAACT 720
 721 TTGGCAAGAGCTGTGAAGGGAGTGACCACTCGGCTCGCTGGAGCTGGAGAGCTGGCC 780
 721 TTGGCAAGAGCTGTGAAGGGAGTGACCACTCGGCTCGCTGGAGCTGGAGAGCTGGCC 780
 781 GAGCTGGTGGCTCAGTGGCTCTTGTGGAGGTCGCTGGCTCCCAACCAAGAGCTG 840
 781 GAGCTGGTGGCTCAGTGGCTCTTGTGGAGGTCGCTGGCTCCCAACCAAGAGCTG 840
 841 CTGCCCTGTGAGATGCGCTGCAATGAGAGCTGGGCAAGTCTGTGGTGGCCAAAGTAA 900
 841 CTGCCCTGTGAGATGCGCTGCAATGAGAGCTGGGCAAGTCTGTGGTGGCCAAAGTAA 900
 901 ATTCGGAGAGCACCATTCTTAACATGAGCATGCTCACCGTGAAGCTGGGTGAGCCAAA 960
 901 ATTCGGAGAGCACCATTCTTAACATGAGCATGCTCACCGTGAAGCTGGGTGAGCCAAA 960
 961 GCCATCTCTCTGAAGCATCTTTAACTTGTGGGCAAGAGTCTGTGCTGCTGCTGCTG 1020
 961 GCCATCTCTCTGAAGCATCTTTAACTTGTGGGCAAGAGTCTGTGCTGCTGCTGCTG 1020
 1021 GAGGATGACACCATCATGGAAGATTTGGTAGATAATCATGCAAAAAAATCAAGTCTTAA 1080
 1021 GAGGATGACACCATCATGGAAGATTTGGTAGATAATCATGCAAAAAAATCAAGTCTTAA 1080

RESULT 3
 AAL53993
 ID AAL53993 standard; DNA; 1380 BP.
 XX
 AC AAL53993;
 XX
 DT 06-MAR-2003 (first entry)
 XX
 DE DNA encoding a human SA synthetase protein.
 XX
 KW Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
 KW GlcNAc-2; epimerase; UDP-GlcNAc; mannosyl; (Man)Nac; sialic acid;
 KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
 KW transporter; sialylated glycoprotein; human; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT CDS 1..1080
 FT /tag= a
 FT /product= "Human SA synthetase protein"
 PN US2002142386-A1.
 XX
 PD 03-OCT-2002.
 XX
 PF 15-AUG-2001; 2001US-00930440.
 XX
 PR 02-MAR-1999; 99US-0122582P.
 PR 08-DEC-1999; 99US-0169624P.
 PR 25-AUG-2000; 2000US-0227579P.
 XX
 (BETE/) BETENBAUGH M J.
 (LAWR/) LAWRENCE S.
 (LEEV/) LEE Y C.
 (COLE/) COLEMAN T A.

PS Disclosure; Page 103-105; 144pp; English.

CC The present sequence is that of human cDNA encoding sialic acid
 CC synthetase (see AY96101), an enzyme that condenses Mann6-P or Man-6-P
 CC with PP6 to form Neu5Ac and KDN phosphates, respectively. This novel
 CC gene, termed SAS, was identified on the basis of homology with the
 CC Escherichia coli neuB gene using a bioinformatics based approach.
 CC Northern blots indicated ubiquitous transcription of the human synthetase
 CC gene in a selection of tissues. The invention provides methods and
 CC recombinantly engineered cells for producing glycoproteins having
 CC sialylated oligosaccharides. The methods involve altering the expression
 CC of enzymes involved in carbohydrate processing. A claimed cell producing
 CC sialylated glycoprotein at above endogenous levels expresses at least 1
 CC (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme
 CC catalyzing the conversion of UDP-GlcNAc to ManNAc, sialic acid
 CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid
 CC transporter at above endogenous levels. Endogenous N-
 CC acetylglucosaminidase activity may be suppressed. A claimed method for
 CC manipulating glycoprotein in an insect cell comprises enhancing the
 CC expression of 1 of the above enzymes, and a claimed method for producing
 CC sialylated glycoproteins involves expressing a heterologous protein
 CC (especially plasminogen, transferrin, Nat,Ki-Arase or thyrotropin) in
 CC the insect cell. Yeast, insect, fungal, plant and bacterial host cells
 CC can be engineered to produce new forms of sialylated glycoproteins,
 CC higher concentrations of sialylated glycoproteins and/or elevated
 CC concentrations of donor substrates (e.g. nucleotide sugars) required for
 CC sialylation

XX SQ Sequence 1146 BP; 290 A; 280 C; 332 G; 244 T; 0 U; 0 Other;

Query Match 100.0%; Score 1080; DB 3; Length 1146;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGTGGGGGGGCAACACCCGCTTC 60
 DB ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGTGGGGGGGCAACACCCGCTTC 105
 QY 61 ATCATTCGCGAGTCGCGAGACACACAGGCGGACCTGGAGCTAGCCAGCGCATGATC 120
 DB ATCATTCGCGAGTCGCGAGACACACAGGCGGACCTGGAGCTAGCCAGCGCATGATC 165
 QY 121 CGCATGGCCAGGAGTGTGGGCTGATTGTCCAAAGTTCGAGAGAGTGAATTC 180
 DB CGCATGGCCAGGAGTGTGGGCTGATTGTCCAAAGTTCGAGAGAGTGAATTC 225
 QY 181 AAGTTTAAATCGAAAGCTTGGAGAGGCGATACACCTCGAAGCATTCCTGGGGGAAGCG 240
 DB AAGTTTAAATCGAAAGCTTGGAGAGGCGATACACCTCGAAGCATTCCTGGGGGAAGCG 285
 QY 241 TACGGGAGCACACACGACATCTCGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 300
 DB TACGGGAGCACACACGACATCTCGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 345
 QY 301 TACCCGAGGAGGTGGGATCTTCTTCACTCCCTCTGGCATGGATGAGATGGCAGTTGAA 360
 DB TACCCGAGGAGGTGGGATCTTCTTCACTCCCTCTGGCATGGATGAGATGGCAGTTGAA 405
 QY 361 TTCTGTGATGAATGATGTTCCATTTTCAAGTGGATCTGGAGACACTAATAATTTT 420
 DB TTCTGTGATGAATGATGTTCCATTTTCAAGTGGATCTGGAGACACTAATAATTTT 465
 QY 421 CCTTATCTGAAAGAGACGCAAAAAGTCCGCCAATGGTGATCTCCAGTGGGATCGAG 480
 DB CCTTATCTGAAAGAGACGCAAAAAGTCCGCCAATGGTGATCTCCAGTGGGATCGAG 525
 QY 481 TCAATGACACCATGAAGAGAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGC 540
 DB TCAATGACACCATGAAGAGAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGC 585
 QY 541 TTCTTGAGTGTACAGCGCATCCCGCTCCAGCTCAGGACGTCACACCTCGCGGTGATC 600
 DB TTCTTGAGTGTACAGCGCATCCCGCTCCAGCTCAGGACGTCACACCTCGCGGTGATC 645

QY 601 TCGGAATATCAGAAGCTCTTTCTGACATTCCTCATAGGGTATTTCTGGGCATGAACAGGC 660
 DB TCGGAATATCAGAAGCTCTTTCTGACATTCCTCATAGGGTATTTCTGGGCATGAACAGGC 705
 QY 661 ATAGCGATATCTGTGGCCCGCAGTGGCTCTGGGGGGCCAAAGTGTGTACACAACT 720
 DB ATAGCGATATCTGTGGCCCGCAGTGGCTCTGGGGGGCCAAAGTGTGTGTGAACGTCACTAACT 765
 QY 721 TTGGCAACAGACCTGGAAGGGGAGTGACCACTCGGCCTCGCTGGAGCTTGAGAACTGGCC 780
 DB TTGGCAACAGACCTGGAAGGGGAGTGACCACTCGGCCTCGCTGGAGCTTGAGAACTGGCC 825
 QY 781 GAGCTGGTGGCGTGAAGTGGCTCTTTGTGGAGCGTGGCCCTGGGCTCCCCAACCCAGCAGCTG 840
 DB GAGCTGGTGGCGTGAAGTGGCTCTTTGTGGAGCGTGGCCCTGGGCTCCCCAACCCAGCAGCTG 885
 QY 841 CTGGCCCTGTGAGATGGCCCTGCAATGAGAGCTGGGCAACTCTGTGTGGCCAAAGTGAA 900
 DB CTGGCCCTGTGAGATGGCCCTGCAATGAGAGCTGGGCAACTCTGTGTGGCCAAAGTGAA 945
 QY 901 ATTCGGAGGCGCACCATTCCTAAATGAGCATGCTCACCGTCAAGTGGTGGAGCCCAA 960
 DB ATTCGGAGGCGCACCATTCCTAAATGAGCATGCTCACCGTCAAGTGGTGGAGCCCAA 1005
 QY 961 GCCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGGTCCTGGTCACTGTGAA 1020
 DB GCCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGGTCCTGGTCACTGTGAA 1065
 QY 1021 GAGGATGACCATCATCTGGAAGATTTGGTAGATAATCATGGCAAAAATCAAGTCTTAA 1080
 DB GAGGATGACCATCATCTGGAAGATTTGGTAGATAATCATGGCAAAAATCAAGTCTTAA 1125

RESULT 5

ABL90630

ID ABL90630 standard; cDNA; 1268 BP.

AC ABL90630;

DT 24-MAY-2002 (first entry)

DE Human polynucleotide SEQ ID NO 1192.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 XX neurological disease; infection; human; secreted protein; gene; ss.

OS Homo sapiens.

PN WC200190304-A2.

PD 29-NOV-2001.

PF 18-MAY-2001; 2001WO-US016450.

PR 19-MAY-2000; 2000US-0205515P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

DR P-PSDB; ABB90221.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 XX disorders.

PS Claim 4; SEQ ID NO 1192; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins (AB89047-AB89044) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

XX
SQ Sequence 1268 BP; 328 A; 304 C; 364 G; 258 T; 0 U; 14 Other;

Query Match 99.8%; Score 1078; DS 6; Length 1268;
Best Local Similarity 99.8%; Pred. NC. 2.1e-314;
Matches 1078; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCCGCTGGAGCTGGAGCTGTCTCCCGGGCGGTGGGTGGGGCGCAACACCGTGCTTC 60
Db 121 ATCCCGCTGGAGCTGGAGCTGTCTCCCGGGCGGTGGGTGGGGCGCAACACCGTGCTTC 180
QY 61 ATCATTTGCCAGATCGCGCAGAACACACACAGGGGCGA CTTGGACGTAGCCAGACGATGATC 120
Db 181 ATCATTTGCCAGATCGCGCAGAACACACACAGGGGCGA CTTGGACGTAGCCAGACGATGATC 240
QY 121 CGCATGGCAAGAGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 180
Db 241 CGCATGGCAAGAGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 300
QY 181 AAGTTTAAATCGGAAGACCTTTGGAGAGCCATACACCTCGAAGCATTCCTGGGGGAGAGC 240
Db 301 AAGTTTAAATCGGAAGACCTTTGGAGAGCCATACACCTCGAAGCATTCCTGGGGGAGAGC 360
QY 241 TACGGGGAGCAAAACAGCATCTGGAGTTTCCAGCATGACGACGACGAGGAGCTGCAGAG 300
Db 361 TACGGGGAGCAAAACAGCATCTGGAGTTTCCAGCATGACGACGACGAGGAGCTGCAGAG 420
QY 301 TACGGGGAGCAAAACAGCATCTGGAGTTTCCAGCATGACGACGACGAGGAGCTGCAGAG 360
Db 421 TACGGGGAGCAAAACAGCATCTGGAGTTTCCAGCATGACGACGACGAGGAGCTGCAGAG 480
QY 361 TTCTTGATGAATGAATGTTTCAATTTTCAAGTTGGATCTGGAGACACTAATAATTTT 420
Db 481 TTCTTGATGAATGAATGTTTCAATTTTCAAGTTGGATCTGGAGACACTAATAATTTT 540
QY 421 CCTTATCTGGAAAGACAGCCAAAGAGTGGCGCAATGGTGTATCTCCAGTGGGATGCAG 480
Db 541 CCTTATCTGGAAAGACAGCCAAAGAGTGGCGCAATGGTGTATCTCCAGTGGGATGCAG 600
QY 481 TCAATGGACACCATGAAGCAGTTTATCAGATCTGTGAAGCCCTCAACCCCACTTCGCG 540
Db 601 TCAATGGACACCATGAAGCAGTTTATCAGATCTGTGAAGCCCTCAACCCCACTTCGCG 660
QY 541 TTCTTGATGAATGAATGTTTCAATTTTCAAGTTGGATCTGGAGACACTAATAATTTT 600
Db 661 TTCTTGATGAATGAATGTTTCAATTTTCAAGTTGGATCTGGAGACACTAATAATTTT 720
QY 601 TCGGAATATCAAGAGCTTTCTGACATCTCCATAGGGTATCTGGGATATGAACAGGC 660
Db 721 TCGGAATATCAAGAGCTTTCTGACATCTCCATAGGGTATCTGGGATATGAACAGGC 780
QY 661 ATAGCGATATCTGGCGCCAGTGGCTCTGGGGGCGAGGTGTGGAAGCTCACATACT 720
Db 781 ATAGCGATATCTGGCGCCAGTGGCTCTGGGGGCGAGGTGTGGAAGCTCACATACT 840

QY 721 TTGGACAAGACCTTGGAAAGGGAGTGACCACTCGGCGCTCGCTGGAGACCTGGAGAACTGGCC 780
Db 841 TTGGACAAGACCTTGGAAAGGGAGTGACCACTCGGCGCTCGCTGGAGACCTGGAGAACTGGCC 900
QY 781 GAGCTGTGGCGGTCAAGTGGCTCTTGTGGAGCGTCCCTGGGCTCCCAACCAAGCAGCTG 840
Db 901 GAGCTGTGGCGGTCAAGTGGCTCTTGTGGAGCGTCCCTGGGCTCCCAACCAAGCAGCTG 960
QY 841 CTGCCCTGTGAGATGGCTGCATGAGAAAGCTGGGCAAGTCTGTGGTGGCCCAAGTGA 900
Db 961 CTGCCCTGTGAGATGGCTGCATGAGAAAGCTGGGCAAGTCTGTGGTGGCCCAAGTGA 1020
QY 901 ATTCCGGAGGACCACTTCTTAAATGAGATGATGCTACCGTGAAGTGGGTGGTGGCCCAA 960
Db 1021 ATTCCGGAGGACCACTTCTTAAATGAGATGATGCTACCGTGAAGTGGGTGGTGGCCCAA 1080
QY 961 GCCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGGTCACTGTTGAA 1020
Db 1081 GCCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGGTCACTGTTGAA 1140
QY 1021 GAGATGACACCATCATCTGGAAGATTTGGTAGATAATCATGGCAAAAAATCAAGTCTTAA 1080
Db 1141 GAGATGACACCATCATCTGGAAGATTTGGTAGATAATCATGGCAAAAAATCAAGTCTTAA 1200

RESULT 6
AA159142
ID AA159142 standard; cDNA; 1230 BP.
XX
AC AA159142;
XX
DC 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1345.

XX Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

OS Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX 23-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00662191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.

(HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Q;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR P-PSDB; AAM39986.
XX

PT Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.

PS Claim 1; SEQ ID NO 1345; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA158642-AA162213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities, chemotactic/chemokinetic activity, haemostatic
CC Activin/inhibin activity, cancer diagnosis and therapy, drug screening,
CC and thrombotic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 1230 BP; 310 A; 309 C; 358 G; 253 T; 0 U; 0 Other;

Query Match 99.7%; Score 1076.8; DB 4; Length 1230;
Best Local Similarity 99.8%; Pred. No. 6.2e-314;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCGCTGGAGCTGGAGCTGTGTCCTCCGGCGCTGGTGGGGCGGCGGACACCCGCTGCTC 60
DB |||||
QY 61 ATCATTCGCGAGATCGGCGCAGACACCCAGCGCGACCTGGAGCTAGCCAGCGCATGATC 120
DB |||||
QY 179 ATCATTCGCGAGATCGGCGCAGACACCCAGCGCGACCTGGAGCTAGCCAGCGCATGATC 238
DB |||||
QY 121 CGCATGCGCGAGAGTGTGGGGCTGATGTCGCAAGTTCGAGAGAGTGTGATGATC 180
DB |||||
QY 239 CGCATGCGCGAGAGTGTGGGGCTGATGTCGCAAGTTCGAGAGAGTGTGATGATC 298
DB |||||
QY 181 AAGTTTAAATCGGAAGCTTGGAGAGCGCATACACCTCGAGAGCTTCTGGGGGAGAGCG 240
DB |||||
QY 299 AAGTTTAAATCGGAAGCTTGGAGAGCGCATACACCTCGAGAGCTTCTGGGGGAGAGCG 358
DB |||||
QY 241 TAGCGGGAGCACAACGACATCTGGAGTTCAGCCATGACACAGTACAGGAGCTGCAGAGG 300
DB |||||
QY 359 TAGCGGGAGCACAACGACATCTGGAGTTCAGCCATGACACAGTACAGGAGCTGCAGAGG 418
DB |||||
QY 301 TAGCGGGAGAGTGTGGATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB |||||
QY 419 TAGCGGGAGAGTGTGGATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 478
DB |||||
QY 361 TTCTGCTGATGAACCTGAATGTTTCCATTTTCAAGTGTGATCTGAGACACACTAATATTT 420
DB |||||
QY 479 TTCTGCTGATGAACCTGAATGTTTCCATTTTCAAGTGTGATCTGAGACACACTAATATTT 538
DB |||||
QY 421 CTTTATCTGGAAGAAGCAGGCAAAAGGTCGCCAAATGCTGATCTCCAGTGGGATGCGAG 480
DB |||||
QY 539 CTTTATCTGGAAGAAGCAGGCAAAAGGTCGCCAAATGCTGATCTCCAGTGGGATGCGAG 598
DB |||||
QY 481 TCAATGACACCATGAGAGAGTTTATCAGTCTGTGAGCCCTCAACCCCACTTCTGCG 540
DB |||||
QY 599 TCAATGACACCATGAGAGAGTTTATCAGTCTGTGAGCCCTCAACCCCACTTCTGCG 658
DB |||||
QY 541 TTCTTGAGTGTACAGCGCATACCCCTCCAGCTGAGGAGCTCAACCTGCGGGTCTATC 600
DB |||||
QY 659 TTCTTGAGTGTACAGCGCATACCCCTCCAGCTGAGGAGCTCAACCTGCGGGTCTATC 718
DB |||||
QY 601 TCGAATATCAGAGCTCTTCTTCTGACATTCCTCCATGAGGATTCCTGGCATGAACAGGC 660
DB |||||
QY 719 TCGAATATCAGAGCTCTTCTTCTGACATTCCTCCATGAGGATTCCTGGCATGAACAGGC 778
DB |||||
QY 661 ATAGCGATATCTGTGGCGGAGTGGCTCTGCGGGCCAGAGTGTGAGAGCTCAATAACT 720
DB |||||
QY 779 ATAGCGATATCTGTGGCGGAGTGGCTCTGCGGGCCAGAGTGTGAGAGCTCAATAACT 838
DB |||||
QY 721 TTGACACAGACCTGGAAGGCGAGTGACACTCTGGCTCTGGTGGAGCTGGAGAGCTGGCC 780
DB |||||

DB 839 TTGGACAAAGACCTGGAAGGGAGTGACCACTCGGCTCGCTGAGAGCTGAGAACTGGCC 898
QY 781 GAGCTGGTGGCTCAGTGGCTCTTGTGGAGCTGCTGGCTGGCTGCCCAACCCAGAGAGCTG 840
DB |||||
QY 839 GAGCTGGTGGCTCAGTGGCTCTTGTGGAGCTGCTGGCTGGCTGCCCAACCCAGAGAGCTG 958
DB |||||
QY 841 CTGCGCTGTGAGATGGCTTCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGA 900
DB |||||
QY 959 CTGCGCTGTGAGATGGCTTCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGA 1018
DB |||||
QY 901 ATTCCGGAGGACCACTTCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGA 960
DB |||||
QY 1039 ATTCCGGAGGACCACTTCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTGA 1078
DB |||||
QY 961 GCCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGCTCTGTGTGGTGGCCAAAGTGA 1020
DB |||||
QY 1079 GCCTATCTCTCTGGAAGACATCTTTAATCTAGTGGGCAAGAGCTCTGTGTGGTGGCCAAAGTGA 1138
DB |||||
QY 1021 GAGGATGACACCATCATGGAAGAGTGTGTAGATATCATGTGCAAAAAAATCAAGTCTTAA 1080
DB |||||
QY 1139 GAGGATGACACCATCATGGAAGAGTGTGTAGATATCATGTGCAAAAAAATCAAGTCTTAA 1198
DB |||||
RESULT 7
ADB49125
ID ADB49125 standard; cDNA; 1230 BP.
XX
AC ADB49125;
XX
DT 04-DEC-2003 (first entry)
DE
DE Novel human cDNA SEQ ID NO 1035.
XX
XX ss; cancer; neurodegenerative disease; human.
XX
XX Homo sapiens.
XX
XX US2003104529-A1.
XX
XX 05-JUN-2003.
XX
XX 04-JAN-2002; 2002US-00337270.
XX
XX 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 19-JUL-2000; 2000US-00620312.
XX
XX (ZHOU/) ZHOU P.
PA (TANG/) TANG Y T.
PA (LIUC/) LIU C.
PA (ASUN/) ASUNDI V.
XX (DRMA/) DRMANAC R T.
PI Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;
XX
XX WPI; 2003-678194/64.
XX
XX New polynucleotide, useful for treating diseases e.g., cancer or
PT neurodegenerative diseases.
XX
XX Claim 1; SEQ ID NO 1035; 99pp; English.
XX
CC The invention relates to a polynucleotide comprising a sequence given in
CC the specification, or its mature protein-coding portion, or its
CC complement. The polynucleotide is useful for treating diseases e.g.,
CC cancer or neurodegenerative diseases and many others listed in the
CC specification. The present sequence represents a novel human cDNA. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=20030104529.
XX
XX Sequence 1230 BP; 310 A; 309 C; 358 G; 253 T; 0 U; 0 Other;

QY 1 ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGGTGGCGGGCAACACCCGCTGCTTC 60
Db 71 ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGGTGGCGGGCAACACCCGCTGCTTC 130
QY 61 ATCATTCGCGAGATCGCGCAAGAACACACAGCGGACCTGGACGTAGCCAAAGCGCATGTC 120
Db 131 ATCATTCGCGAGATCGCGCAAGAACACACAGCGGACCTGGACGTAGCCAAAGCGCATGTC 190
QY 121 CGCATGCCAAGGAGTGGGGCTGATTGTGCCAAGTTCACAGAGAGTGGAGTCAAAATTC 180
Db 191 CGCATGCCAAGGAGTGGGGCTGATTGTGCCAAGTTCACAGAGAGTGGAGTCAAAATTC 250
QY 181 AAGTTTAAATCGAAAGCTTCGAGAGGACATACACCTCGAAGCAATCTCTGGGGGAAGACG 240
Db 251 AAGTTTAAATCGAAAGCTTCGAGAGGACATACACCTCGAAGCAATCTCTGGGGGAAGACG 310
QY 241 TACGGGAGACAAACGACATCTGGAGTTCAGCCATGACACAGGAGTCGAGAGG 300
Db 311 TACGGGAGACAAACGACATCTGGAGTTCAGCCATGACACAGGAGTCGAGAGG 370
QY 301 TACGCCGAGGAGTGGGATCTTCTTCACTGCGCTCTGSCATGGATGAGATGCGAGTTCGA 360
Db 371 TACGCCGAGGAGTGGGATCTTCTTCACTGCGCTCTGSCATGGATGAGATGCGAGTTCGA 430
QY 361 TTCTGATGAATGAATGTTCCATTTTAAAGTTGATCTGAGAGACACTAAATATTT 420
Db 431 TTCTGATGAATGAATGTTCCATTTTAAAGTTGATCTGAGAGACACTAAATATTT 490
QY 421 CCTTATCTGGAAGAGACAGCCAAAGGTGCGCCCAATGGTGATCTCCAGTGGAGTCAG 480
Db 491 CCTTATCTGGAAGAGACAGCCAAAGGTGCGCCCAATGGTGATCTCCAGTGGAGTCAG 550
QY 481 TCATGAGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 540
Db 551 TCATGAGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 610
QY 541 TTCTGAGTGTACAGGAGATGACATCCGCTCCAGCTGAGGACGTCAACCTCGGGTCACT 600
Db 611 TTCTGAGTGTACAGGAGATGACATCCGCTCCAGCTGAGGACGTCAACCTCGGGTCACT 670
QY 601 TCGAATATCAGAGCTCTTCTCGATTCCTGAGATCCCATAGGTATCTGGCATGAACAGGC 660
Db 671 TCGAATATCAGAGCTCTTCTCGATTCCTGAGATCCCATAGGTATCTGGCATGAACAGGC 730
QY 661 ATAGCGATATCTGTGGCGCAGTGGCTCTGGGGCCCAAGGTGTTTGAACGTCACTAACT 720
Db 731 ATAGCGATATCTGTGGCGCAGTGGCTCTGGGGCCCAAGGTGTTTGAACGTCACTAACT 790
QY 721 TTGGACAGACCTGGAAGGGAGTGAACATCGGCTCGCTGGAGCTCGTGGAGAACTGGCC 780
Db 791 TTGGACAGACCTGGAAGGGAGTGAACATCGGCTCGCTGGAGCTCGTGGAGAACTGGCC 850
QY 781 GAGCTGTGCGGTGAGTGGCTGTGTGGAGGTGCTGCTGGGCTCCCAACCAAGCAGCTG 840
Db 851 GAGCTGTGCGGTGAGTGGCTGTGTGGAGGTGCTGCTGGGCTCCCAACCAAGCAGCTG 910
QY 841 CTGCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTCAAA 900
Db 911 CTGCTGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGTGGTGGCCAAAGTCAAA 970
QY 901 ATTCCGGAAGGACCATCTTAAACATGAGATGCTACCGTGAAGGTGGGTGAGCCAAA 960
Db 971 ATTCCGGAAGGACCATCTTAAACATGAGATGCTACCGTGAAGGTGGGTGAGCCAAA 1030
QY 961 GCCTATCTCTGAGACATCTTAACTAGTGGGCAAGGTCTCTGCTCACTGTTGAA 1020
Db 1031 GCCTATCTCTGAGACATCTTAACTAGTGGGCAAGGTCTCTGCTCACTGTTGAA 1090
QY 1021 GAGGATGACACCATCATGGAAGATTTGGTAGAATATCATGGCAAAAAAATCAAGTCTTAA 1080
Db 1091 GAGGATGACACCATCATGGAAGATTTGGTAGAATATCATGGCAAAAAAATCAAGTCTTAA 1150

RESULT 9
AAI60928
ID AAI60928 standard; cDNA; 1243 BP.
XX AAI60928;
AC AAI60928;
XX 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4917.
XX Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.
XX Homo sapiens.
OS WC200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US014263.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR P-PSDB; AAM41772.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
as central nervous system injuries.
Claim 1; SEQ ID NO 4917; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
encoded polypeptides (AAM38642-AAM42213) with notropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX Sequence 1243 BP; 298 A; 291 C; 385 G; 269 T; 0 U; 0 Other;
SQ

Query Match 83.2%; Score 898.2; DB 4; Length 1243;
Best Local Similarity 88.6%; Pred. No. 4.5e-260;
Matches 1041; Conservative 0; Mismatches 3; Indels 131; Gaps 1;
QY 1 ATGCGCTGGAGCTGGAGCTGTGTCCCGGGCGCTGGGTGGCGGGCAACACCCGCTGCTTC 60

Db 69 ATGCCGCTGGAGCTGGAGCTGTCCCGGGGCTGGTGGGGGCAACACCCGCTGCTC 128
QY 61 ATCATTTCCGAGATCGGCCAGAACCAACAGGGGACCTGGACGTACCCAAAGCGCATGATC 120
Db 129 ATCATTTCCGAGATCGGCCAGAACCAACAGGGGACCTGGATGTACCCAAAGCGCATGATC 188
QY 121 CCATGTCGCAAGAGTGTGGGCTGATCTGCCAGTTCACAGAGAGTGTAGTAATTC 180
Db 189 CGATGGCCAGAGGTGTGGGCTGATGTGCCAGTTCACAGAGAGTGTAGTAATTC 248
QY 181 AAGTTTAATCGAAAGCTTTGGAGAGCCATACACCTCGAAGCAATTCCTGGGGAGAGCG 240
Db 249 AAGTTTAATCGAAAGCTTTGGAGAGCCATACACCTCGAAGCAATTCCTGGGGAGAGCG 308
QY 241 TAGGGGAGACAAAGACATCTGGAGTTTCAGGCATTCAGTACAGGAGCTGCAGAG 300
Db 309 TAGGGGAGACAAAGACATCTGGAGTTTCAGGCATTCAGTACAGGAGCTGCAGAG 368
QY 301 TAGCCGAGAGGCTTGGGATCTTCTCACTGCTCTGSCATGATCAGATGGCAGTTGAA 360
Db 369 TAGCCGAGAGGCTTGGGATCTTCTCACTGCTCTGSCATGATCAGATGGCAGTTGAA 428
QY 361 TTCCTGCATSAATGATCTTCATTTTCAAGTTTSGATCTGGAGACACTAATAATTT 420
Db 429 TTCTGCATSAATGATCTTCATTTTCAAGTTTSGATCTGGAGACACTAATAATTT 468
QY 421 CCTTATCTGMAAAGACAGCAAAA-----446
Db 489 CCTTATCTGMAAAGACAGCAAAAAGGGCTGGCACAGTGTTTAAGAGATGTCGTG 548
QY 447 -----446
Db 549 GAGTTCAGTTGATGATGATGAGACGACAGTGTGGATGTGTGGGAGAGTCAGAACCTCTA 608
QY 447 -----AGTCCGCCAATGGTGATCTCA 469
Db 609 AAGAGAAAGTGTGATGTGTGGTGTCTGTGATTAATCACTGAGTCCCAATGGTGATCTCA 668
QY 470 GTGGATGAGTCAATGGACACCACTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAAC 529
Db 669 GTGGATGAGTCAATGGACACCACTGAAGCAAGTTTATCAGATCGTGAAGCCCTCAAC 728
QY 530 CCAACTCTGCTTCTGCAAGTGTACAGGCGATACCGGCTCCAGCTCAGGAGCTCAAC 589
Db 729 CCAACTCTGCTTCTGCAAGTGTACAGGCGATACCGGCTCCAGCTCAGGAGCTCAAC 788
QY 590 TCGGGTCACTCGGAATACAGAGCTCTTCTCAGATCCCATTCAGAGGATTTCTGGGC 649
Db 789 TCGGGTCACTCGGAATACAGAGCTCTTCTCAGATCCCATTCAGAGGATTTCTGGGC 848
QY 650 ATGAAACAGGATAGCGATCTGTGGCGCAGTGGCTCTGGGGCCAGAGGTGTGGAC 709
Db 849 ATGAAACAGGATAGCGATCTGTGGCGCAGTGGCTCTGGGGCCAGAGGTGTGGAC 808
QY 710 GTCACTAATCTTGAACAAGCTTGAAGGAGTGCACCTCGGCTCGCTGGAGCTG 769
Db 909 GTCACTAATCTTGAACAAGCTTGAAGGAGTGCACCTCGGCTCGCTGGAGCTG 968
QY 770 GAGACTCGCGAGCTGTGGGTGAGTGTGAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGG 829
Db 965 GAGACTCGCGAGCTGTGGGTGAGTGTGAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGG 1028
QY 830 CCAAGCAGCTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 889
Db 1029 CCAAGCAGCTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1088
QY 890 CCAAGTGAATTTCCGAGGACCAATTTCAACATGACATGCTTCACCTGAGAGCTGG 949
Db 1089 CCAAGTGAATTTCCGAGGACCAATTTCAACATGACATGCTTCACCTGAGAGCTGG 1148
QY 950 GTGAGCCCAAGCTTATCTCTGAGAGCACTTTTAATCTAGTGGGCAAGAGGCTCTGG 1009
Db 1149 GTGAGCCCAAGCTTATCTCTGAGAGCACTTTTAATCTAGTGGGCAAGAGGCTCTGG 1208

QY 1010 TCACCTGTTGAAGAGATGACACCATCATCGAAGAA 1044
Db 1209 TCACCTGTTGAAGAGATGACACCATCATCGAAGAA 1243
RESULT 10
AAS33141/C
ID AAS33141 standard; cDNA; 1238 BP.
XX
AC AAS33141;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seq ID No 100.
XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ss.
XX Homo sapiens.
OS
PN NC200155326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001347.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 09-SEP-2000; 2000US-02311242P.
 PR 09-SEP-2000; 2000US-02311243P.
 PR 08-SEP-2000; 2000US-02311244P.
 PR 08-SEP-2000; 2000US-02311433P.
 PR 08-SEP-2000; 2000US-02311434P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 27-SEP-2000; 2000US-0236327P.
 PR 28-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0239355P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246539P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-451931/48.
 XX P-PSDB; AAU20432.
 DR New nucleic acids and polypeptides, useful for diagnosing, preventing or
 XX treating medical conditions.
 PT Claim 1; SEQ ID NO 100; 753pp; English.
 XX The invention relates to novel isolated nucleic acid molecules (I)
 CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardiac-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and
 CC thrombosis), infections caused by bacteria, viruses and fungi and ocular
 CC disorders (e.g. corneal infections). (I) and (II), agonists, antagonists
 CC and antibodies can also be used to promote wound healing, maintain organs
 CC before transplantation, and support cell culture of primary tissues.
 Query Match 67.8%; Score 732; DB 4; Length 1238;
 Best Local Similarity 99.7%; Pred. No. 6.5e-210;
 Matches 732; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 347 AGATGGCAGTTGAATTCCTGCGAAGTGAATGTTCCATTTCATTAAGTGGATCGGAG 406
 Db 799 AGATGGCAGTTGAATTCCTGCGAAGTGAATGTTCCATTTCATTAAGTGGATCGGAG 740
 Qy 407 ACACTATATTTTCCTTCTGAAAGACAGCAAAAGGTCGCCATGTGATCT 466

739 ACACATAATATTTTCTTATCTGGAAGAGACAGCCAAAAGGTCGCCCAATGGTGTCT 680
 467 CCAGTGGAGATGAGTCATATGACACATGACACAGTATTTATCAGATCGTGAAGCCCTCA 526
 679 CCAGTGGAGATGAGTCATATGACACATGACACAGTATTTATCAGATCGTGAAGCCCTCA 620
 527 ACCCCAACTTCTGCTTCTTGGAGTACACAGGCGATACCCGCTCCAGCCTGAGGACGTCA 586
 619 ACCCCAACTTCTGCTTCTTGGAGTACACAGGCGATACCCGCTCCAGCCTGAGGACGTCA 560
 587 ACCTCGGGTCTATCTCGAATATCAGAACTCTTTCTTGACATCCCATFAGGATTTCTG 646
 559 ACCTCGGGTCTATCTCGAATATCAGAACTCTTTCTTGACATCCCATFAGGATTTCTG 500
 647 GGCATGAACAGGATAGGATATCTGTCGCGAGTGGCTCTGGGGGCGAGGTTGG 706
 499 GGCATGAACAGGATAGGATATCTGTCGCGAGTGGCTCTGGGGGCGAGGTTGG 440
 707 AACGTCACATACTTTGACAAAGACTGGAAGGGAGTGACCACTGGGCTCGCTGGAGC 766
 439 AACGTCACATACTTTGACAAAGACTGGAAGGGAGTGACCACTGGGCTCGCTGGAGC 380
 767 CTGGAGACTGCGGAGTGTGGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 826
 379 CTGGAGACTGCGGAGTGTGGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 320
 827 CAACCAAGAGTGTGGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 886
 319 CAACCAAGAGTGTGGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 260
 887 TGGCCAAAGTGAATTCGGAGGACCACTTCTTAACTGACATGCTCACCGTGAAGG 946
 259 TGGCCAAAGTGAATTCGGAGGACCACTTCTTAACTGACATGCTCACCGTGAAGG 200
 947 TGGTGAAGCCCAAGCTTATCTCTCTGGAAGATCTTTAATCTAGTGGGCAAGAGTCC 1036
 199 TGGTGAAGCCCAAGCTTATCTCTCTGGAAGATCTTTAATCTAGTGGGCAAGAGTCC 140
 1007 TGGTCACTGTTGAAGAGGATGACACATCATGGAAGATTTGGTAGATATCATGGCAAA 1066
 139 TGGTCACTGTTGAAGAGGATGACACATCATGGAAGATTTGGTAGATATCATGGCAAA 80
 1067 AAATCAAGTCTTAA 1080
 79 AATCAAGTCTTAA 66
 RESULT 11
 ID AAH06847
 XX standard; cDNA; 813 BP.
 AC AAH06847;
 DT 26-JUN-2001 (first entry)
 XX Human cDNA clone (5'-primer) SEQ ID NO:3682.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-00116126.
 XX 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-0030253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.
 Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 WPI; 2001-318749/34.
 Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 Claim 1; SEQ ID NO 3682; 2537pp + Sequence Listing; English.
 The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the 5602 complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.
 Sequence 813 BP; 198 A; 199 C; 243 G; 168 T; 0 U; 5 Other;
 Query Match 65.8%; Score 710.2; DB 4; Length 813;
 Best Local Similarity 98.7%; Pred. NO. 2e-203;
 Matches 734; Conservative 0; Mismatches 8; Indels 2; Gaps 2;
 QY 1 ATGCCGCTGGAGCTGAGCTGTGTCCCGGGCGCTGGTGGGGGCGGCAACACCGTGTCTTC 60
 DB 71 ATGCCGCTGGAGCTGAGCTGTGTCCCGGGCGCTGGTGGGGGCGGCAACACCGTGTCTTC 130
 QY 61 ATCATTTGCCGAGATGCGCCAGAACACACAGGGGCGCTGGAGCTGAGCTAGCAATTC 120
 DB 131 ATCATTTGCCGAGATGCGCCAGAACACACAGGGGCGCTGGAGCTGAGCTAGCAATTC 190
 QY 121 CGCATGGCCCAAGAGTGTGGGCTGATTTCTGCCAGTTCACAGAGAGTGTAGCAATTC 180
 DB 191 CGCATGGCCCAAGAGTGTGGGCTGATTTCTGCCAGTTCACAGAGAGTGTAGCAATTC 250
 QY 181 AAGTTTAATCGGAAAGCCCTTGGAGAGGCCATACACCTCGAAGCATTCCTGGGGGAAGAG 240
 DB 251 AAGTTTAATCGGAAAGCCCTTGGAGAGGCCATACACCTCGAAGCATTCCTGGGGGAAGAG 310
 QY 241 TAGGGGAGACAAACAGCATCTGGAGTTCAGGCTACGCTACAGGAGGCTGCGAGG 300
 DB 311 TAGGGGAGACAAACAGCATCTGGAGTTCAGGCTACGCTACAGGAGGCTGCGAGG 370
 QY 301 TAGCGCGAGGAGTGGGATCTTCTTCACTGCTCTGGCATGTGATGGAGTGGAGTTGAA 360
 DB 371 TAGCGCGAGGAGTGGGATCTTCTTCACTGCTCTGGCATGTGATGGAGTGGAGTTGAA 430
 QY 361 TTCTCTGCATGAACCTGAATGTTCCATTTTCAAAGTTGGATCTGGAGACACTAATATTTT 420
 DB 431 TTCTCTGCATGAACCTGAATGTTCCATTTTCAAAGTTGGATCTGGAGACACTAATATTTT 490
 QY 421 CCTTATCTGGAAAAGACAGCCAAAAGGTGCGCCCATGCTGTATCTCCAGTGGATGCGAG 480

Db 491 CCTTCTCTGAAAGACAGNCAAAAAAGGTGCGCCCAATGGTGTGATCTCCAGTGGATCAG 550
 QY 481 TCAATGACACCATGAGAGCAAGTTTATCATGATGTTGAGGCCCTCAACCCCAACTTTCG 540
 Db 551 TCAATGACACCATGAGAGCAAGTTTATCATGATGTTGAGGCCCTCAACCCCAACTTTCG 610
 QY 541 TTCTTCAGTGTACCAAGCGCATACCGCTCCAGCTGAGGACGTCACACCTGGGGTATC 600
 Db 611 TTCTTCAGTGTACCAAGCGCATACCGCTCCAGCTGAGGACGTCACACCTGGGGTATC 670
 QY 601 TCGCAATATCAGAAAGCTCTTCTCTGACATCCCATAGGATTTCTGGGCATGAACAGGC 660
 Db 671 TCGCAATATCAGAAAGCTCTTCTCTGACATCCCATAGGATTTCTGGGCATGAACAGGC 730
 QY 661 ATAGCGATACTGTGGCCGAGTGGC-TCTGGGGGCCAAGTG-TGAAGCTGCATTAAC 719
 Db 731 ATAGCGATACTGTGGCCGAGTGGC-TCTGGGGGCCAAGTG-TGAAGCTGCATTAAC 789
 QY 720 TTTGGACACACCTGGAAGGGAG 743
 Db 790 TTTGGACACACCTGGAAGGGAG 813

RESULT 12

AAH11876/C

ID AAH11876 standard; cDNA; 514 BP.

XX AAH11876;

XX AC

XX DT

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (3'-primer) SEQ ID NO:8711.

XX DE

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS

XX Homo sapiens.

XX XX

XX EP1074617-A2.

XX XX

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX XX

XX (HELI-) HELIX RES INST.

XX PA

XX PT

XX Ora T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX XX

XX WPI; 2001-3-8749/34.

XX XX

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX XX

XX Claim 3; SEQ ID NO 8711; 2537pp + Sequence Listing; English.

XX PS

XX The present invention describes primer sets for synthesising 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dT primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX cDNAs.

CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX XX

SQ Sequence 514 BP; 112 A; 149 C; 113 G; 133 T; 0 U; 7 Other;

Query Match

Best Local Similarity 44.8%; Score 483.8; DB 4; Length 514;

Matches 485; Conservative 98.2%; Pred. No. 3.3e-135;

Mismatches 485; Indels 9; Gaps 0;

QY 587 ACCTGCGGTCATCTCGGAATATCAGAAAGCTCTTCTCCAGATCCCATAGGTTATCTG 546
 Db 514 ACNTGNGGGTNTNTNGGAANATCAGAAAGCTCTTCTCCAGATCCCATAGGTTATCTG 455
 QY 647 GGCATGAAACAGGCATAGCGATATCTGTGCGCGAGTGGCTCTGGGGCCCAAGGTGTTGG 706
 Db 454 GGCATGAAACAGGCATAGCGATATCTGTGCGCGAGTGGCTCTGGGGCCCAAGGTGTTGG 395
 QY 707 AACCTCATAAATCTTTGGACAAGACCTTGGAAAGGGAGTGACCACTCGGCTCTCGTGGAGC 766
 Db 394 AACCTCATAAATCTTTGGACAAGACCTTGGAAAGGGAGTGACCACTCGGCTCTCGTGGAGC 335
 QY 767 CTGCGAAGCTGGCGAGCTGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 826
 Db 334 CTGCGAAGCTGGCGAGCTGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 275
 QY 827 CAACCAAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 886
 Db 274 CAACCAAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 215
 QY 987 TGGCCAAAGTGAAATTCGGGAAGGACCACTTCTTAAACAATGGAATGCTCCCGTGAAGG 946
 Db 214 TGGCCAAAGTGAAATTCGGGAAGGACCACTTCTTAAACAATGGAATGCTCCCGTGAAGG 155
 QY 947 TGGTGTAGCCCAAGCCCTATCTCTGAGACATCTTAAATCTAGTGGGCAAGAGGTCC 1006
 Db 154 TGGTGTAGCCCAAGCCCTATCTCTGAGACATCTTAAATCTAGTGGGCAAGAGGTCC 95
 QY 1007 TGGTCACTGTGAGAGGATGACACCATCATGGAGAATTTGGTAGATAATCATGGCAAAA 1066
 Db 94 TGGTCACTGTGAGAGGATGACACCATCATGGAGAATTTGGTAGATAATCATGGCAAAA 35
 QY 1067 AAATCAAGTCTTAA 1080
 Db 34 AAATCAAGTCTTAA 21

RESULT 13

AAS91837

ID AAS91837 standard; cDNA; 1417 BP.

XX AAS91837;

XX AC

XX AAS91837;

XX DT

XX 13-FEB-2002 (first entry)

XX XX

XX DNA encoding novel human diagnostic protein #27641.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS

XX Homo sapiens.

XX XX

XX WO200175067-A2.

XX XX

PD 11-OCT-2001.
 XX 30-MAR-2002; 2001WO-US008631.
 PF 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649157.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX P-PSDB; ABQ27650.
 DR New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 1; SEQ ID NO 27641; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1417 BP; 353 A; 336 C; 409 G; 319 T; 0 U; 0 Other;
 Query Match 39.5%; Score 426.4; DB 5; Length 1417;
 Best Local Similarity 89.4%; Pred. No. 1.2e-117;
 Matches 588; Conservative 0; Mismatches 46; Indels 24; Gaps 11;
 QY 447 AGGTGCGCCCAATGGTGAATCTCCAGTGGGATGCAATGACACCATGAAGCAAGTTTA 506
 DB 427 AGGTGCGCCCAATGGTGAATCTCCAGTGGGATGCAATGACACCATGAAGCAAGTTAT 486
 QY 507 TCAGATGCTGAAGCCCTCAACCCCAATCTTCGCTTCCTTTCAGTGTACCGCGCATACC 565
 DB 487 TCAAAATGCTGAAGCCCTCAACCCCAATCTTCGCTTCCTTTCAGTGTACCGCGCATACC 546
 QY 566 CGCTCCAGCTGAGACCTCAACC--TCGGGTGATCTCGGAATATCAGAGCTCTTTCC 623
 DB 547 CGCTCCAGCTGAGACCTCAACCCTCGGGTCACTCGGAATATCAGAGCTCTTTCC 606
 QY 624 TGACATTTCCC---ATAGAGGTATCTTGGGCATGAACAGCGCATAGGATATCTGT---GGC 677
 DB 607 TGACATTTCCCATTAGGAGTATCTTGGGCATGAACAGCGCATAGGATATCTTTGGCC 666
 QY 678 CGCAGTGGCTTGGGG--CCAGGTGTGGAA-----CGTACATAACTTTGGCAAGAC 731
 DB 667 CCCAGTGGCTTGGGGGCCCAAGGTGTGGAAAGCTCTCACTAAACTTTGGCAAGAC 726
 QY 732 CTGGAAGGGGAGTG---ACCACTGGGCTCGCT--GGAGCTGGAGAACTGGCGG-AGC 784
 DB 727 CGGGAGAGGGGAGTGTGACCCACTCGGCTCGCTTGGAGCCCTGGAGAACTGGCGGAAAGC 786

QY 785 TGGTGGGTGAGTGGTCTTGTGGAGCGTGCCTGGGTCTCCCAACC--AAGCAGCTGCTG 843
 DB 787 TGGTGGGTGAGTGGTCTTGTGGAGCGTGCCTGGGTCTCCCAACCCTAGCAGCTGCTG 846
 QY 844 CCTGTGAGATGGCTGCAA--TGAGAACTGGGCAAGTCTGTGGTGGCCAAAGTGAATAAT 902
 DB 847 CCTGTGAGATGGCTGCAANTGAACTGGGCAAGTCTGTGGTGGCCAAAGTGAATAAT 906
 QY 903 TCCGGAAGGCCACCATTTCAACATGACATGCTCACTCCGCTGAGAGTGGGTGAGCCCAAGC 962
 DB 907 TCCGGAAGGCCACCATTTCAACATGACATGCTCACTCCGCTGAGAGTGGGTGAGCCCAAGC 966
 QY 963 CTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGCTCACTGTTGAGA 1022
 DB 967 CTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGCTCACTGTTGAGA 1026
 QY 1023 GGATGACACCATCATGGAAGAAATTCGTAGATTAATCATGGCAAPAAAATCAAGTCTTAA 1080
 DB 1027 GGATGACACCATCATGGAAGAAATTCGTAGATTAATCATGGCAAPAAAATCAAGTCTTAA 1084
 RESULT 14
 AEN95016/c
 ID AEN95016 standard; DNA; 464 BP.
 XX AEN95016;
 AC AEN95016;
 DT 13-AUG-2002 (first entry)
 XX Gene #1514 used to diagnose liver cancer.
 DE Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumour; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX Homo sapiens.
 OS WC200229103-A2.
 PN 11-APR-2002.
 PD 02-OCT-2001; 2001WO-US30589.
 PF 02-OCT-2000; 2000US-0237054P.
 PR (GENE-) GENE LOGIC INC.
 PA Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 XX WE1; 2002-426119/45.
 DR Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.
 XX Claim 1; SEQ ID NO 1514; 298pp; English.
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumour in a patient, and differentiating metastatic liver cancer from of
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 464 BP; 101 A; 133 C; 103 G; 127 T; 0 U; 0 Other;
 Query Match 38.7%; Score 418; DB 6; Length 464;
 Best Local Similarity 98.6%; P-val. No. 2.2e-115;
 Matches 432; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
 XX
 CC 643 TCTGGCCATGAACAGGCGATAGCGATATCTGTGGCCGCGAGTGCTTGGGGGCGAAGTG 702
 CC 644 TCTGGCCATGAACAGGCGATAGCGATATCTGTGGCCGCGAGTGCTTGGGGGCGAAGTG 703
 CC 703 TTTGGAACGTCACATACTTTGACAAAGACTGGAGGGGAGTGACCACTCGGCTCGCTG 762
 CC 404 TTGGACGTCACATACTTTGACAAAGACTGGAGGGGAGTGACCACTCGGCTCGCTG 345
 CC 763 GAGCCTGGAGAACTGGCCGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 822
 CC 344 GAGCCTGGAGAACTGGCCGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 285
 CC 823 TCCCAACCAAGCAGCTGTGCTGCTGTGAGTGGCTGCAATGAGAACTGGGCAAGTCT 882
 CC 284 TCCCAACCAAGCAGCTGTGCTGCTGTGAGTGGCTGCAATGAGAACTGGGCAAGTCT 225
 CC 883 GTGGTGGCCAAAGTGAATTTCCGGAAGGACCACTTTCAATGAGAACTGGTCAAGCTG 942
 CC 224 GTGGTGGCCAAAGTGAATTTCCGGAAGGACCACTTTCAATGAGAACTGGTCAAGCTG 165
 CC 943 AGGTGGGTGAGCCCAAGCCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAG 1002
 CC 164 AGGTGGGTGAGCCCAAGCCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAG 106
 CC 1003 GTCTGGTCACTGTTGAAGAGGATGACACCATCATCGAAGAAATTTGATAGATATCGC 1062
 CC 105 GTCTGGTCACTGTTGAAGAGGATGACACCATCATCGAAGAAATTTGATAGATATCGC 46
 CC 1063 AAAAAATCAAGTCTTAA 1080
 CC 45 AAAAAATCAAGTCTTAA 28
 RESULT 15
 ABK94731/c
 ID ABK94731 standard; DNA; 553 BP.
 XX
 AC ABK94731;
 XX
 DT 28-AUG-2002 (first entry)
 XX
 DE Neurodegenerative disease associated polynucleotide #40.
 XX
 KW Mitochondrial function; Forkhead Related Activator-2; FREAC-2; apoptosis;
 KW neurodegenerative disease; proliferative disorder; cancer;
 KW neurological disease; Alzheimer's disease; AD; multiple sclerosis; MS;
 KW Parkinson's disease; PD; vascular dementia; diffuse Lewy body disease;
 KW DLB; amyotrophic lateral sclerosis; ALS;
 KW myoclonic epilepsy lactic acidosis and stroke; MELAS; MERRF;
 KW myoclonic epilepsy ragged red fibre syndrome; autoimmune disease;
 KW Gene therapy; antisense therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0202240996-A2.
 XX
 PD 23-MAY-2002.
 XX
 PF 08-NOV-2001; 2001W0-US046699.
 XX
 PR 08-NOV-2000; 2001S-00709754.
 XX
 PA (MITO-) MITOKOR.
 XX
 PI Hwang J, Becker DK, Davis RE;
 XX
 RR WPI; 2002-508336/54.
 XX

GenCore version: 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 ; Search time 3122.76 Seconds
(without alignments)

10327.779 Million cell updates/sec

Title: US-09-930-440B-5

Perfect score: 1080

Sequence: 1 atgcgcgtgagctggagct.....gcaaaaaaatcaagtccttaa 1080

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931030276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estinu.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_esti.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_est6.*

15: em_estom.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_pbg.*

27: em_gss_vri.*

28: gb_gssi.*

29: gb_gse2.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1076.8	99.7	1080	29 AY414837	AY414837 Homo sapi
C 2	952.6	88.2	984	9 AL578658	AL578658 AL578658
C 3	949.6	87.9	201	13 BX384733	BX384733 BX384733
C 4	917.8	85.0	990	9 AL578579	AL578579 AL578579

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	883.2	81.8	1080	29	AY414837	1080 bp DNA linear	GSS 17-DEC-2003
6	883.2	81.8	1906	11	AK076290	Homo sapiens NANS gene, VIRTUAL TRANSCR-PT, partial sequence,	
7	870.6	80.6	2109	11	AK088912	Genomic survey sequence.	
8	856.6	79.3	925	12	BI517350	AY414837	
9	856.4	79.3	987	9	AL556387	AY414837.1	GI:39770796
10	843	78.1	1005	9	AL556278	GSS.	
11	840	77.8	1201	9	AL515470	Homo sapiens (human)	
12	830	76.9	989	9	AL526241	Homo sapiens	
13	823.4	76.2	889	13	BU187658	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
14	803.2	74.4	872	13	BQ432284	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
15	796.4	73.7	1031	13	BQ064521	1 (bases 1 to 1080)	
16	795.8	73.7	842	9	AL522851	Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejarawal,A.,	
17	792.2	73.4	1049	12	BM541313	Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,	
18	789.4	73.1	948	29	AY414838	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,	
19	786	72.8	1046	9	AL526281	Adams,M.D. and Cargill,M.	
20	784.4	72.6	890	13	BU181010	Inferring nonneutral evolution from human-chimp-mouse orthologous	
21	775	71.8	979	12	BG681730	gene trilos	
22	772	71.5	905	13	BQ893527	Science 302 (5652), 1960-1963 (2003)	
23	764.4	70.8	1097	12	BM806099	14671302	
24	757.4	70.1	878	9	AL522852	2 (bases 1 to 1080)	
25	754.4	69.9	878	12	BI824205	Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejarawal,A.,	
26	749.2	69.4	1070	9	AL544111	Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,	
27	737.6	68.3	1050	12	BM563430	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,	
28	736.2	68.2	773	12	BI517963	Adams,M.D. and Cargill,M.	
29	735.4	68.1	845	13	BU528624	Inferring nonneutral evolution from human-chimp-mouse orthologous	
30	726.2	67.2	905	13	BQ649507	gene trilos	
31	722.2	66.9	883	12	BI258475	Science 302 (5652), 1960-1963 (2003)	
32	721.6	66.8	777	12	BG568099	14671302	
33	717.4	66.4	978	12	BG114903	2 (bases 1 to 1080)	
34	710.6	65.8	919	13	BU195376	Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejarawal,A.,	
35	710.2	65.8	813	9	AU132828	Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,	
36	703.2	65.1	1031	9	AL544197	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,	
37	695	64.4	706	12	BG746292	Adams,M.D. and Cargill,M.	
38	690	63.9	690	12	BG746329	Inferring nonneutral evolution from human-chimp-mouse orthologous	
39	689.4	63.8	835	10	BE614630	gene trilos	
40	687.8	63.7	948	10	BE793764	Science 302 (5652), 1960-1963 (2003)	
41	686.4	63.6	769	13	BU618375	14671302	
42	682.2	63.2	779	14	CA417480	2 (bases 1 to 1080)	
43	680.4	63.0	931	12	BI525002	Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejarawal,A.,	
44	678.4	62.8	844	13	BU167760	Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,	
45	675.2	62.5	877	12	BI905036	Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,	

ALIGNMENTS

RESULT 1

AY414837

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES

source Location/Qualifiers

gene 1..1080

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

<..>.1080

/gene="NANS"

/locus_tag="HCM5357"

ORIGIN

Query Match 99.7%; Score 1076.8; DB 29; Length 1080;

Best Local Similarity 99.8%; Pred. No. 3.4e-268;

Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGCAACACCCGTGCTTC 60

DB 1 ATCCCGCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGCAACACCCGTGCTTC 60

QY 61 ATCATTTGCCAGATTCGGCCAGACACACACAGGCGGACCTGGAGCTAGCCAGCGCATGATC 120

DB 61 ATCATTTGCCAGATTCGGCCAGACACACAGGCGGACCTGGAGCTAGCCAGCGCATGATC 120

QY 121 CGCATGGCCCAAGAGTGTGGGGCTGATTGTGCCAAGTTCCAGAGAGTGTAGATTTC 180

DB 121 CGCATGGCCCAAGAGTGTGGGGCTGATTGTGCCAAGTTCCAGAGAGTGTAGATTTC 180

QY 181 AAGTTTAATCGGAAGCTTTGGAGGCGCCATACACCTCGAAGCATTCCTGGGGGAAGCG 240

DB 181 AAGTTTAATCGGAAGCTTTGGAGGCGCCATACACCTCGAAGCATTCCTGGGGGAAGCG 240

QY 241 TAGCGGGAGCACAAAGACATCTGGAGTTTCAGCATGACACAGTACAGGAGCTGCAGG 300

DB 241 TAGCGGGAGCACAAAGACATCTGGAGTTTCAGCATGACACAGTACAGGAGCTGCAGG 300

QY 301 TAGCCGAGAGGTGTGGATCTTTCTTCACGTCCCTCGGCATGATGATGATGGAGTTGAA 360

DB 301 TAGCCGAGAGGTGTGGATCTTTCTTCACGTCCCTCGGCATGATGATGATGGAGTTGAA 360

QY 361 TTCTTCATGAATGTAATGTTCCATTTTCAAAGTTGATCTCGAGACATTAATAATTTT 420

DB 361 TTCTTCATGAATGTAATGTTCCATTTTCAAAGTTGATCTCGAGACATTAATAATTTT 420

QY 421 CCTTATCTGAAAGACAGCCAAAGAGTCCGCAATGATGATCTCCAGTGGGATGCAG 480

DB 421 CCTTATCTGAAAGACAGCCAAAGAGTCCGCAATGATGATCTCCAGTGGGATGCAG 480

QY 481 TCAATGGACACCATGAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 540

DB 481 TCAATGGACACCATGAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGC 540

QY 541 TTCTTCAGTGTACGACGCAATACCCGCTCGAGCTGAGGACGTCAACCTGGGGTCAATC 600

DB 541 TTCTTCAGTGTACGACGCAATACCCGCTCGAGCTGAGGACGTCAACCTGGGGTCAATC 600

QY 601 TCGGAATATCAGAGCTCTTCTCTGACATTCATCCATAGGTTATCTGGGATGAAACAGGC 660

DB 601 TCGGAATATCAGAGCTCTTCTCTGACATTCATCCATAGGTTATCTGGGATGAAACAGGC 660

QY 661 ATAGCGATATCTGTGGCGCAGTGGCTCTGSGGGCCCAAGGTGTGGAACTCAATAACT 720

DB 661 ATAGCGATATCTGTGGCGCAGTGGCTCTGSGGGCCCAAGGTGTGGAACTCAATAACT 720

QY 721 TTGGAACAAGCTTGGAGGGAGTGAACCACTGGGCTCTGTGGAGCTTGAGAACTGGCC 780

DB 721 TTGGAACAAGCTTGGAGGGAGTGAACCACTGGGCTCTGTGGAGCTTGAGAACTGGCC 780

QY 781 GAGCTGTGGGTGAGTGGCTCTGTGGAGCTGGCTCTGGCTCCGCCCAACAGGAGCTG 840

DB 781 GAGCTGTGGGTGAGTGGCTCTGTGGAGCTGGCTCTGGCTCCGCCCAACAGGAGCTG 840

QY 841 CTGCCCTGTGAGTGCCTGCAATGAGAAGCTGCGCAAGTCTGTGTGGGCGCAAGTGA 900

DB 841 CTGCCCTGTGAGTGCCTGCAATGAGAAGCTGCGCAAGTCTGTGTGGGCGCAAGTGA 900

QY 901 ATTCCGGAAGGACCACTTTCAACAATGGACATGCTACCGTGAAGTGGGTGAGCCAAA 960

DB 901 ATTCCGGAAGGACCACTTTCAACAATGGACATGCTACCGTGAAGTGGGTGAGCCAAA 960

QY 961 GCCTATCTCTCCGCAAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGGTCACTGTTGAA 1020

DB 961 GCCTATCTCTCTCCGCAAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGGTCACTGTTGAA 1020

QY 1021 GAGGTGACACCACTCATGGAAGATTTGTTAGATAATCATGGCAAAAAAATCAAGTCTTAA 1080

DB 1021 GAGGTGACACCACTCATGGAAGATTTGTTAGATAATCATGGCAAAAAAATCAAGTCTTAA 1080

RESULT 2

AL578658/c

LOCUS

DEFINITION

AL578658 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens cDNA clone CS0DK004Y114 3-PRIME, mRNA sequence.

ACCESSION

AL578658

VERSION

AL578658.2

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 984)

AUTHORS

Li, W.B., Gruber, C., Jesse, J., and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On Feb 16, 2001 this sequence version replaced gi:12942945.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqres@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5047.f for more information about this cluster, see

http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DK004BE07NP1&cluster=5047.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope, sequence ID : CS0DK004BE07NP1.

FEATURES

Location/Qualifiers

1..984

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK004Y114"

/cell_type="HELA CELLS COT 25-NORMALIZED"

/clone_line="HELA"

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/notes="First strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 88.2%; Score 952.6; DB 9; Length 984;

Best Local Similarity 98.4%; Pred. No. 5.7e-236;

Matches 966; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

QY 99 GGCCTGAGCCAGCGCATGATCGCATGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATG 158

DB 984 GGCCTGAGCCAGCGCATGATCGCATGCGCATGCGCATGCGCATGCGCATGCGCATG 925

QY 159 CCAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 215

DB 924 CCAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 865


```

QY 541 TTCTTCAGTGTACCGGACACACCGCTCCAGCTGAGGAGCTCAACCTGCGGCTCATC 600
Db |||||||
QY 529 TTCTTCAGTGTACCGGACACACCGCTCCAGCTGAGGAGCTCAA-CTGCGGCTCATC 471
Db |||||||
QY 601 TCGGAATATCAGAGCTCTTTCTCTGATCATCCATAGGTPATTTCTGGGCACTGAACAGGC 660
Db |||||||
QY 470 TCGGAATATCAGAGCTCTTTCTCTGATCATCCATAGGTPATTTCTGGGCACTGAACAGGC 411
Db |||||||
QY 661 ATAGCCATATCTGTGCGCCAGTGTCTCTGGGGCCCAAGTGTGGAACTCATATAACT 720
Db |||||||
QY 410 ATAGCCATATCTGTGCGCCAGTGTCTCTGGGGCCCAAGTGTGGAACTCATATAACT 351
Db |||||||
QY 721 TTGGACAAGACCTGGAAGGAGTGTGACCACTCGGCTCTGCTGAGAGCTTGAGAACTGGCC 780
Db |||||||
QY 350 TTGGACAAGACCTGGAAGGAGTGTGACCACTCGGCTCTGCTGAGAGCTTGAGAACTGGCC 291
Db |||||||
QY 781 GAGCTGGTGGTCACTGGGCTCTTTGGAGAGCTGGCCCTGGGCTGCCCAACCAAGCAGCTG 840
Db |||||||
QY 290 GAGCTGGTGGTCACTGGGCTCTTTGGAGAGCTGGCCCTGGGCTGCCCAACCAAGCAGCTG 231
Db |||||||
QY 841 CTGCCCTGTGAGATGCGCTGCATGAGAGCTGGGCACTGTCTGTGGCCCAAGTGAA 900
Db |||||||
QY 230 CTGCCCTGTGAGATGCGCTGCATGAGAGCTGGGCACTGTCTGTGGCCCAAGTGAA 171
Db |||||||
QY 901 ATTCCGGAGAGGACCACTCTAAACAATGGACAATCTCAACCTGAGAGTGGGTGAGCCCAAA 960
Db |||||||
QY 170 ATTCCGGAGAGGACCACTCTAAACAATGGACAATCTCAACCTGAGAGTGGGTGAGCCCAAA 111
Db |||||||
QY 961 GCTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGTGTCATCTGTGAA 1020
Db |||||||
QY 110 GGTATCTCTCTGAGACATCTTTAATCTAGTGGGCAAGAGTCTCTGTGTCATCTGTGAA 51
Db |||||||
QY 1021 GAGGATGACCACTATCTGGAAGATTTGGTGAATTAATCATGGCAAAAAAT 1070
Db |||||||
QY 50 GAGGATGACCACTATCTGGAAGATTTGGTGAATTAATCATGGCAAAAAAT 1
Db |||||||

```

RESULT 4

```

AL578579/c
LOCUS AL578579 Homo sapiens HELA CELLS COT 25-NORMALIZED EST 01-JUN-2003
DEFINITION cDNA clone CS0DK001YM22 3-PRIME, mRNA sequence.
ACCESSION AL578579
VERSION AL578579.1 GI:12942789
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5047.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK001BG11NP1&cluster=5047.f. Contact :
Feng Liang Email : fliang@life-techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DK001BG11NP1.
Location/Qualifiers
1. .990
/organism="Homo sapiens"
/-cl_type="mRNA"
/-db_xref="taxon:9606"
/clone="CS0DK001YM22"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"

```

FEATURES

source

```

ORIGIN
Query Match 85.0%; Score 917.8; DB 9; Length 990;
Best Local Similarity 96.8%; Pred. No. 6.2e-227; Indels 3; Gaps 3;
Matches 958; Conservative 8; Mismatches 23;
QY 1 ATGCGCGTGGAGCTGGAGCTGTGTCCCGCGTGGTGGTGGCGGCAACCCGTTGCTTC 60
Db |||||||
QY 990 ATGCGCGTGGAGCTGGAGCTTTGTCCCGCGTGGTGGTGGCGGCAACCCGTTGCTTC 931
Db |||||||
QY 61 ATCATTTGCCAGATTCGGCCAGAACCAACAGGCGGACCTGGACGTAGCGAGCGCATATC 120
Db |||||||
QY 930 ATAAATTCOCAGATTCGGCCAGAACCAACAGGCGGACCTGGATGTAGCCAGCGCATATC 871
Db |||||||
QY 121 CGCATGGCCAAAGAGTGTGGGCTGATTGTGCCAAGTTCCAGAGAGTGTAGCTAGAAATC 180
Db |||||||
QY 870 CGCATGGCCAAAGAGTGTGGGCTGATTGTGCCAAGTTCCAGAGAGTGTAGCTAGAAATC 811
Db |||||||
QY 181 AAGCTTTAATCGGAAAGCCTTTGGAGAGGCCATACACCTCGAAGCATTTCTCGGGGAAGACG 242
Db |||||||
QY 810 AAGCTTTAATCGGAAAGCCTTTGGAGAGGCCATACACCTCGAAGCATTTCTCGGGGAAGACG 751
Db |||||||
QY 241 TACGGGGAGCCAAAGACATCTCGAGTTTCAGGCATGACCACTACAGGAGCTGCGACAGG 300
Db |||||||
QY 750 TACGGGGAGCCAAAGACATCTCGAGTTTCAGGCATGACCACTACAGGAGCTGCGACAGG 692
Db |||||||
QY 301 TACGGCGAGAGGTTGGAGTCTTTTCACTGCTCTGGCATGTGATGATGATGCGCATTTGAA 360
Db |||||||
QY 690 TACGGCGAGAGGTTGGAGTCTTTTCACTGCTCTGGCATGTGATGATGATGCGCATTTGGA 632
Db |||||||
QY 361 TTCTTCATGAACTGAAATGTTCATTTTCAAAAGTTGGATCTGGAGACATATAAATTTT 420
Db |||||||
QY 631 TTCTTCATGAACTGAAATGTTCATTTTCAAAAGTTGGATCTGGAGACATATAAATTTT 572
Db |||||||
QY 421 CCTTATCTGGAAAGACAGCCAAAGAGTGGCCCATGTGTATCTCCAGTGGGATCGAG 480
Db |||||||
QY 571 CCTTATCTGGAAAGACAGCCAAAGAGTGGCCCATGTGTATCTCCAGTGGGATCGAG 512
Db |||||||
QY 481 TCAATGGAACACCATGAAGCAAGTTTATCAGATCTGAAGCCCTCAACCCCAACTTCTGC 540
Db |||||||
QY 511 TCAATGGAACACCATGAAGCAAGTTTATCAGATCTGAAGCCCTCAACCCCAACTTCTGC 452
Db |||||||
QY 541 TTCTTCAGGTGACCAAGGATACCGCTCCAGCTGAGGACGTCACCTGGCGGTCATC 600
Db |||||||
QY 451 TTCTTCAGGTGACCAAGGATACCGCTCCAGCTGAGGACGTCACCTGGCGGTCATC 393
Db |||||||
QY 601 TCGGAATATCAGAGCTCTTTCTCTGACATTTCCATAGGTTATTTCTGGGCACTGAGACGC 660
Db |||||||
QY 392 TCGGAATATCAGAGCTCTTTCTCTGACATTTCCATAGGTTATTTCTGGGCACTGAGACGC 333
Db |||||||
QY 661 ATAGCCATATCTGTGCGCCAGTGTCTCTGGGGCCCAAGTGTGGAACTCATATAACT 720
Db |||||||
QY 332 ATAGCCATATCTGTGCGCCAGTGTCTCTGGGGCCCAAGTGTGGAACTCATATAACT 273
Db |||||||
QY 721 TTGGACAAGACCTGGAAGGAGTGTGACCACTCGGCTCTGCTGAGAGCTTGAGAACTGGCC 780
Db |||||||
QY 272 TTGGACAAGACCTGGAAGGAGTGTGACCACTCGGCTCTGCTGAGAGCTTGAGAACTGGCC 213
Db |||||||
QY 781 GAGCTGGTGGTCACTGGGCTCTTTGGAGAGCTGGCCCTGGGCTGCCCAACCAAGCAGCTG 840
Db |||||||
QY 212 GAGCTGGTGGTCACTGGGCTCTTTGGAGAGCTGGCCCTGGGCTGCCCAACCAAGCAGCTG 153
Db |||||||
QY 841 CTGCCCTGTGAGATGCGCTGCATGAGAGCTGGGCACTGTCTGTGGTGGCCAAAGTGAA 900
Db |||||||
QY 152 CTGCCCTGTGAGATGCGCTGCATGAGAGCTGGGCACTGTCTGTGGTGGCCAAAGTGAA 93
Db |||||||
QY 901 ATTCCGGAGAGGACCACTCTAAACAATGGACAATCTCAACCTGAGAGTGGGTGAGC-CCAA 959
Db |||||||

```

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer, five prime end enriched, double-strand cDNA was
 digested with NotI and enriched into the NotI and EcoRV
 sites of the pCMVSPORT 6 vector. Library was normalized."


```

Db      92 ATTCGGGAGGACCAATCTTACAAATGAACATGCTCACCGTGAAGGTGGGTGAGCGCCAA 33
Qy      960 AGCCTATCTCTCTGAAGACATCTTTTAATCTAG 991
Db      32 AGGCTATCTCTCTGAAGACATCTTTTAATCTAG 1

RESULT 5
AY414839      1080 bp DNA linear GSS 17-DEC-2003
LOCUS      Mus musculus NANS gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY414839
VERSION AY414839.1 GI:39770798
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1080)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1080)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source location/Qualifiers
1..1080
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1080
/gene="NANS"
/locus_tag="HCM5357"

Query Match 81.8%; Score 883.2; DB 29; Length 1080;
Best Local Similarity 88.6%; Pred. No. 6.3e-218;
Matches 95; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy      1 ATCCCGCTGGAGCTGGAGCTGTGTCCGGGGCGCTGGGTGGCGGGCCAAACCCCGTGTTC 60
Db      1 ATCCCGCTGGAACTGGAGCTGTGTCCGGGGCGCTGGGTGGGTGGAAAGCACCCGCTGTTC 60
Qy      61 ATCATTCGAGATCGGCGGAGAACCCACAGGCGGACTGGACGTAGCCAGCGCATGATC 120
Db      61 ATCATTCGAGATCGGCGGAGAACCCACAGGAGACATAGATGTGGCCAGCGCATGATC 120
Qy      121 CGCATGCGCAGGAGTGTGGGGCTGTGTGCCAGTTCAGAGAGTGTGAGTGAATTC 180
Db      121 CGCATGCGCAGGAGTGTGGGGCTGTGTGCCAGTTCAGAGAGTGTGAGTGAATTC 180
Qy      181 AAGTTTAAATCGAAAGCTTGGAGAGCCATACACCTCGAAGCATCTCTGGGGGAGACG 240
Db      181 AAGTTTAAATCGAAAGCTTGGAGAGCCATATATCTCGAAGCATCTATGGGGGAGACG 240
Qy      241 TACGGGAGACAAACGACATCTCGAGTTCAGCATGCCATGACCAAGTACAGGAGCTCCAGG 300
Db      241 TATGGGAGACAAACGCGCATCTCGAATTCAGCCAGCACCATGACCAAGGAGCTCCAGG 300
Qy      301 TACGCCGAGGAGGTGGGAATCTTCTCACTGCCTCTGCGATGGATGAGATGGCAGTTGAA 360

```

```

Db      301 TATGGCGAGGAGATCGGCATCTTCTTCACTGCTCTGGCATGACGAGATGCGAGTTGAG 360
Qy      361 TTCCTGCAATGAAGTCAATGTTCCATTTTCAAGTTGGATCTGGAGACACTATAATTTT 420
Db      361 TTCCTGCAATGAAGTCAATGTTCCATTTTCAAGTTGGATCTGGAGACACTATAATTTT 420
Qy      421 CCTTATCTGGAAAGACAGCCAAAGGTGCGCCAAATGGTGTATCTCCAGTGGGATGAG 480
Db      421 CCTTATCTGGAAAGACAGCCAAAGGTGCGCCAAATGGTGTATCTCCAGTGGGATGAG 480
Qy      481 TCATGGGACACCATGAAGCAAGTTTATCAGATCTGTAAGCCCTCTCAACCCCACTTCTGC 540
Db      481 TCATGGGACACCATGAAGCAAGTTTATCAGATCTGTAAGCCCTCTCAACCCCACTTCTGC 540
Qy      541 TTCCTGAGTGTACCCAGGCGCATACCCGCTCCAGCCCTGAGGAGCTCAACCTGGGGTCAATC 600
Db      541 TTCCTGAGTGTACCCAGGCGCATACCCGCTCCAGCCCTGAGGAGCTCAACCTGGGGTCAATC 600
Qy      601 TCGGAATATCAGAGCTCTTTCCTGCAATTCCTCATAGGATTTCTGGGATGAACAGGC 660
Db      601 TCGGAATATCAGAGCTCTTTCCTGCAATTCCTCATAGGATTTCTGGGATGAACAGGC 660
Qy      661 ATAGCGATATCTGGGCGCGCAGTGGCTCTGGGGGCGCAAGTGTGTGAAGCTCACATAACT 720
Db      661 ATAGCGATATCTGGGCGCGCAGTGGCTCTGGGGGCGCAAGTGTGTGAAGCTCACATAACT 720
Qy      721 TTGGACAAGACTCTGGAAGGGAGTGAACCACTCGGCTCTGCTGAGGCTGAGACTGGGC 780
Db      721 TTGGACAAGACTCTGGAAGGGAGTGAACCACTCGGCTCTGCTGAGGCTGAGACTGGGC 780
Qy      781 GAGCTGGTGGGTGAGTGGCTCTGCTGAGGCTGAGCTGGCTGGCTGGCTGGCTGGCTGGCT 840
Db      781 GAGCTGGTGGGTGAGTGGCTCTGCTGAGGCTGAGCTGGCTGGCTGGCTGGCTGGCTGGCT 840
Qy      841 CTGCGCTGTGAGTGGCTGCAATGAGAGCTGGGCGAGTCTGTGTGGTGGCCAAAGTGA 900
Db      841 CTGCGCTGTGAGTGGCTGCAATGAGAGCTGGGCGAGTCTGTGTGGTGGCCAAAGTGA 900
Qy      901 ATTCGGAAGCCACCATCTTAACATGACATGCTCACCGTGAAGTGGTGGTGGCCAA 960
Db      901 ATTCGGAAGCCACCATCTTAACATGACATGCTCACCGTGAAGTGGTGGTGGCCAA 960
Qy      961 GCCTATCTCTCGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGCTCACTGTGAA 1020
Db      961 GCCTATCTCTCGAAGACATCTTTAATCTAGTGGGCAAGAGGTCTCTGCTCACTGTGAA 1020
Qy      1021 GAGATGACACCATCATGGAAGATTTGATATATCATGCGCAAAAATCAAGTCTTAA 1080
Db      1021 GAGATGACACCATCATGGAAGATTTGATATATCATGCGCAAAAATCAAGTCTTAA 1080

RESULT 6
AK076290      1906 bp mRNA linear HTC 20-SEP-2003
LOCUS      Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched
DEFINITION library, clone:4632418E04 product:sialic acid synthase, full insert
sequence.
ACCESSION AK076290
VERSION AK076290.1 GI:26345279
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

```

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 204939374
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kusunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, K., Kishine, T., Harada, A.,
 Yamamoto, R., Matsunaga, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 1306)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
 Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
 Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
 Tejima, Y., Toyota, T., Yamamura, T., Yamazaki, I., Yasunishi, A.,
 Yoshida, K., Yoshino, N., Muramatsu, M. and Hayashizaki, Y.
 Direct Submision
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/.
 Location/Qualifiers
 1. 1906
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:4632418E04"
 /db_xref="MGI:1903268"
 /db_xref="taxon:10090"
 /clone="4632418E04"
 /tissue_type="skin"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="0 day neonate"
 32. 1111
 /note="unnamed protein product; putative"

sialic acid synthase (MGD|G1:2149820, GB|NM_053179,
 evidence: BLASTN, 100%, match=1140)"
 /codon_start=1
 /protein_id="BAC36290.1"
 /db_xref="GI:26345280"
 /translation="MPLELELCPRWVGKHPFCIIAETGQNEQGGIDVAKRMRTAK
 EGADCAKPKSELEKFKNRKALERDPTKSGNDTKYGEHKLHEESHOYKELOSYA
 QSIGIFFTASGDMEMAVEFLHNLNVPFFKVGSGDTNNPYLEKTAAGKHPMTISSMG
 SMDTKQVYQIVKPLNPFCFLOCTISAYPILOPEDANLRLVISEYKLFPPDIPISYSH
 TGLAISAAVALGAKVLERHTIDTKTKGSDSHSIESPGLAEIYSVRLVERALGSP
 TKOLLPCENACNEKLGKSVAVKVPAGTTLTLDMLTVKVGEPKGPPEPDIENLAGKK
 VLVITIEDDTVNEESVESHSKKKA"
 ORIGIN
 Query Match 81.8%; Score 883.2; DB 11; Length 1306;
 Best Local Similarity 88.6%; Pred. No. 8.5e-218; Indels 0; Gaps 0;
 Matches 957; Conservative 0; Mismatches 123;
 1 ATGCCCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGCAACACCCGTGCTTC 60
 32 ATGCCCTGGAGCTGGAGCTGTGTCCCGGCGCTGGGTGGCGGGCAACACCCGTGCTTC 91
 61 ATCATTCGAGATCGGCCAGAACACACCGGGCGACCTGGACCTAGCAGACGACGATGATC 120
 92 ATCATTCGAGATCGGCCAGAACACACCGGGCGACCTAGATGATGCGGACGATGATC 151
 121 CGCATGGCCAGAGGTGTGGGCTGATTCGCCAGTTCAGAGAGAGTGGAGTATTC 180
 152 CGCATGGCCAGAGGTGTGGGCGGACTGCGCTAGTTTCAGAGAGAGTGGAGTATTC 211
 181 AAGTTTAAATCGGAAAGCCCTTGGAGAGCCATACACCTCGAAGCATTCCTGGGGAAGACG 240
 212 AAGTTTAAATCGGAAAGCCCTTGGAGAGCCATATATCTTGGAGCATTCATGGGGAAGACG 271
 241 TAGGGGAGACACAAAGCATCTGGAGTTGACCATACAGTACAGGAGCTGCGAGG 300
 272 TAGGGGAGACACAAAGCATCTGGAAATTCAGCACACGACGATGAGGAGCTGCGAGG 331
 301 TAGGGGAGAGGTGGGATCTCTTCACTGCTCTGGCATGATGATGAGGAGTGA 360
 332 TAGGGGAGAGATGGGATCTCTTCACTGCTCTGGCATGATGAGGAGTGA 391
 361 TCCCTGCATGAATGAATTCATTTTCAAGTTGGATCTGGAGACACTAATAATTTT 420
 392 TCTCTGCAGAACTGAATGTTTCCCTTTTCAAGTTGGATCTGGAGACACTAATAATTTT 451
 421 CTTATCTGGAAAAGACAGACGACGACGACGATGCTGATCTCGATGAGGATGCGAG 480
 452 CCTTACCTGGAAAAGACAGACGACGACGATGCTGATCTCGATGAGGATGCGAG 511
 481 TCAATGGACACCATGAAGCAAGTTCATCAGATCGTGAAGCCCTCAACCCCACTTCG 540
 512 TCATGGACACCATGAAGCAAGTTCATCAGATCGTGAAGCCCTCAACCCCACTTCG 571
 541 TCTTCTGAGTACACAGGACATACCGGTCGAGCTGAGGAGTGAAGTCACTCGGGGTCATC 600
 572 TCTCTCAATGACACAGGCGGTACCCATACAGCCGAGGATGCCAACCTCGGCGTCATC 631
 601 TCGGATATCAGAGCTCTTTCCTGACATTCCTCCATAGGATTCCTGGGATGAAACAGGC 660
 632 TCGGATATCAGAGCTCTTTCCTGACATTCCTCCATAGGATTCCTGGGATGAAACAGGC 691
 661 ATAGCGATATCTGTGGCGCGAGTGGCTCTGGGGGGCAAGGTGTGGAAAGTCACTAATCT 720
 692 ATCGGCATATCTGTGGCGCGAGTGGCTCTGGGGGGCAAGGTGTGGAAAGTCACTAATCT 751
 721 TTGGACAAAGCTGGAAAGGAGTGAACCACTCGGCTCTGCTGGAGCTGAGAACTGGGCC 780
 752 TTGGACAAAGCTGGAAAGGAGTGAACCACTCAGCTCTGCTGGAGCTGAGAACTGGGCC 811
 781 GAGCTGGTGGCTGAGTGGCTGTGGAGAGTGGCTGGGCTGCCCAACCAAGAGCTG 840
 812 GAGCTGGTGGCTGAGTGGCTGTGGAGAGTGGGCTGGGCTGCCCAACCAAGAGCTG 871

Qy 841 CTGCCCTGTGAGATGGCTGCAATGAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAA 900
 Db 872 CTGCCCTGTGAGATGGCTGCAATGAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAA 931
 Qy 901 ATTCGGAGGACCAATCTTAACATGACATGCTCACCGTGAGAGTGGTGGAGCCCAAA 960
 Db 932 ATCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 991
 Qy 961 GCCTATCTCTCTGAAGCATCTTAACTAGTGGCAAGAGGCTCTGGTCACTGTTCGA 1020
 Db 992 GGCTATCTCTCTGAAGCATCTTAACTAGTGGCAAGAGGCTCTGGTCACTATCGA 1051
 Qy 1021 GAGGATGACACCATCATGGAAGAATGGTAGATAATCATGCGCAAAAATAAATCAAGTCTTAA 1080
 Db 1052 GAAGATGACACCGGTCTATGGAAGAATCCGTGGAAAGTACACGCAAGAAATCAAGGCTTAA 1111

RESULT 7
 AK088912
 LOCUS Mus musculus 2:09 bp mRNA linear HTC 20-SEP-2003
 DEFINITION full-length enriched library, clone:E430030P04 product:sialic acid synthase, full insert sequence.

ACCESSION
 VERSION AK088912.1 GI:26104987
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
 1 Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,H., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499574
 PUBMED 11042159

REFERENCE
 3 Shibata,K., Itoh,H., Aizawa,K., Nagao,K., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,X., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Shikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RTKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplex capillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5

REFERENCE
 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 2109)
 JOURNAL Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,

Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome-gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL:http://genome-gsc.riken.go.jp/
 URL:http://fantom-gsc.riken.go.jp/
 Location/Qualifiers
 1..2109
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="PANTOM_DB:E430030P04"
 /db_xref="MGI:2427805"
 /db_xref="taxon:10090"
 /clone="B430030P04"
 /cell_type="thymic cells"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="2 days neonate"
 41..1121
 /note="putative sialic acid synthase (MGD|MGI:2149820, GB|NM_053179, evidence: BLASTN, 100%, match=1140)"

misc_feature
 41..1121
 /note="putative sialic acid synthase (MGD|MGI:2149820, GB|NM_053179, evidence: BLASTN, 100%, match=1140)"

ORIGIN
 Query Match 80.6%; Score 870.6; DB 11; Length 2109;
 Best Local Similarity 88.4%; Pred. No. 1.7e-214;
 Matches 956; Conservative 0; Mismatches 124; Indels 1; Gaps 1;
 Qy 1 ATCCCGCTGGAGCTGGAGCTGTGTCGGCGCGCTGGGTGGCGGCAACACCCGCTGCTTC 60
 Db 41 ATCCCGCTGGAGCTGGAGCTGTGTCGGCGCGCTGGGTGGCGGCAACACCCGCTGCTTC 100
 Qy 61 ATCATTCGCGAGATCGGCCAGAACACCACGAGCGGACCTGGACGTAGCCACCGCATGATC 120
 Db 101 ATCATCGCGAGATCGGCCAGAACACCACGAGGACATAGATGTGGCCACGCGATGATC 160
 Qy 121 CGCATGCCAAGAGTGTGGGGCTGTGATTCGCCAAGTTCGAGAGAGTGGACCTAGATTC 180
 Db 161 CGCACTCCCAAGGAGTGTGGGGCGGACTGGGCTAACTTCAGAGAGGAGCTGGAGTTC 220
 Qy 181 AAGTTTAAATCGGAAGGCTTGGAGAGGCCATACACCTCGAAGCATTCCTGGGGGAAGAGC 240
 Db 221 AAGTTTAAATCGGAAGGCTTGGAGAGGCCATATTAATTCGGAAGCAATTCATGGGGAAGAGC 280
 Qy 241 TAGCGGAGGACCAAAACGACATCTGGAGTTCAGCCATGACCACTACAGGAGCTGCAGAGG 300
 Db 281 TAGCGGAGGACCAAGCGGATCTGGATTTCAGCCACGACCTACAGGAGCTGCAGAGC 340
 Qy 301 TAGCGGAGGAGGTTGGGATCTCTTCACTGGCTCTGGCATGGATGGATGGCATGGATGGAA 360


```

Db      541  AAGACTGGAAGGGGAGTGAACACTCGGCTCGCTGAGAGCTGGAGAACTGGCGAGCTG 600
QY      787  GGGCGGTGAGTGGCTTTGTGAGAGCTGGCTGGGGCTCCCGAACCAAGCAGCTGCTGGCC 846
Db      601  GGGCGGTGAGTGGCTTTGTGAGAGCTGGCTGGGGCTCCCGAACCAAGCAGCTGCTGGCC 660
QY      847  TGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGGTGCCCAAGTGAATTCGG 906
Db      661  TGTGAGATGGCTGCAATGAGAGCTGGGCAAGTCTGTGGTGCCCAAGTGAATTCGG 720
QY      907  GAAGGACCACTTCT-AACAATGACATGCTACCGTGGAAGGTGGGTGAGCCCAAGCCCTA 965
Db      721  GAAGGACCACTTCTAAACAATGACATGCTACCGTGGAAGGTGGGTGAGCCCAAGCCCTA 780
QY      966  TCTCTGAGAGTCTTTAAATCTAGTGGGCAAGAGTCTGTGTGAGAGTCTGTGTGAGAGGA 1025
Db      781  TCTCTGAGAGTCTTTAAATCTAGTGGGCAAGAGTCTGTGTGAGAGTCTGTGTGAGAGGA 840
QY      1026  TGACACCATCATGGAAGATTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1080
Db      841  TGACACCATCATGGAAGATTGTGATGATGATGATGATGATGATGATGATGATGATGATG 895

RESULT 9
LOCUS   AL556387 987 bp mRNA linear EST 31-MAY-2003
DEFINITION AL556387 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK004YI14 5-PRIME, mRNA sequence.
ACCESSION AL556387
VERSION AL556387.2 GI:31278189
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 987)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:12899016.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91056 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5047.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK004BE07QPI&cluster=5047.f. Contact :
Feng Liang E-mail: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DK004BE07QPI.
Location/Qualifiers
1. .987
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK004YI14"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and cloned into the NotI and EcoRV
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 79.3%; Score 856.4; DB 9; Length 987;
Best Local Similarity 99.9%; Pred. No. 5.4e-21;
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCTGGGTGGGGCGGCAACACCCGTGCTTC 60

```

```

Db      130  ATGCCGCTGGAGCTGGAGCTGTGTCTCCGGGGCTGGGTGGGGCGGCAACACCCGTGCTTC 189
QY      61  ATCATTTCCGAGATCGGCCCAAGAACCAACAGGGGCGAAGCTGGACCTAGCAACAGCGCATGATC 120
Db      190  ATCATTTCCGAGATCGGCCCAAGAACCAACAGGGGCGAAGCTGGATGTAGCCAGCGCATGATC 249
QY      121  CGCATGCCCAAGAGTGTGGGGCTGATTGTGCGCAAGTTCAGAGAGTGTAGATTC 180
Db      250  CGCATGCCCAAGAGTGTGGGGCTGATTGTGCGCAAGTTCAGAGAGTGTAGATTC 309
QY      181  AAGTTTAATTCGGAAGACCTTTGGAGGCGCATACACCTCGAAGCATTCCTCGGGGGAACG 240
Db      310  AAGTTTAATTCGGAAGACCTTTGGAGGCGCATACACCTCGAAGCATTCCTCGGGGGAACG 369
QY      241  TAGGGGAGCACAAAACGACATCTGGAGTTCAGCCATGACCAAGTACAGGGAGCTGCAGAGG 300
Db      370  TAGGGGAGCACAAAACGACATCTGGAGTTCAGCCATGACCAAGTACAGGGAGCTGCAGAGG 429
QY      301  TAGGGGAGAGGTGGGATCTTCTTCACTGTGCTCTGGCATGGATGAGATGGCAGTGA 360
Db      430  TAGGGGAGAGGTGGGATCTTCTTCACTGTGCTCTGGCATGGATGAGATGGCAGTGA 489
QY      361  TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
Db      490  TTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 549
QY      421  CTTTATCTGGAAAAGACAGCCAAAGAGTTCGCCCAATGATGATGATGATGATGATGATG 480
Db      550  CTTTATCTGGAAAAGACAGCCAAAGAGTTCGCCCAATGATGATGATGATGATGATGATG 609
QY      481  TCAATGGACACCATGAGAGAGTTCATGATGATGATGATGATGATGATGATGATGATGATG 540
Db      610  TCAATGGACACCATGAGAGAGTTCATGATGATGATGATGATGATGATGATGATGATGATG 669
QY      541  TTCTTGCAGTGTACAGAGCATAACCGCTCCAGCTGAGAGTGTGAGAGTGTGAGAGTGTG 600
Db      670  TTCTTGCAGTGTACAGAGCATAACCGCTCCAGCTGAGAGTGTGAGAGTGTGAGAGTGTG 729
QY      601  TCGAATATCAGAGTCTTTCTTGCATGATGATGATGATGATGATGATGATGATGATGATG 660
Db      730  TCGAATATCAGAGTCTTTCTTGCATGATGATGATGATGATGATGATGATGATGATGATG 789
QY      661  ATAGCGATATCTGTGGCGCAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG 720
Db      790  ATAGCGATATCTGTGGCGCAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG 849
QY      721  TTGACACAGACCTGGAAGGGAGTGCACCATCGGCTCGCTGAGGCTGTGAGAGTGTGAG 780
Db      850  TTGACACAGACCTGGAAGGGAGTGCACCATCGGCTCGCTGAGGCTGTGAGAGTGTGAG 909
QY      781  GAGCTGTGGCTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG 840
Db      910  GAGCTGTGGCTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG 969
QY      841  CTGCCCTGTGAGATGGCC 858
Db      970  CTGCCCTGTGAGATGGCC 987

RESULT 10
LOCUS   AL556278 1005 bp mRNA linear EST 31-MAY-2003
DEFINITION AL556278 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK001YM22 5-PRIME, mRNA sequence.
ACCESSION AL556278
VERSION AL556278.2 GI:31278081
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1005)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

```


QY 61 ATCATTTGCCGAGATCGGCCAGAACACACAGGCGGACCTGGACGTAGCCAAAGCGCATGATC 120
 Db 163 ATCATTTGCCGAGATCGGCCAGAACACACAGGCGGACCTGGATGTAGCCAAAGCGCATGATC 222
 QY 121 CGCATGGCCAGGAGCTGGGGCTGATTTGTGCCAGTTCACAGAGAGTGAGCTAGATTC 180
 Db 223 CGCATGGCCAGGAGCTGGGGCTGATTTGTGCCAGTTCACAGAGAGTGAGCTAGATTC 282
 QY 181 AAGTTTAAATCGAAAGCCTTGGAGAGGCGCATACACTCGAACATTCCTGGGGGAAGAGC 240
 Db 283 AAGTTTAAATCGAAAGCCTTGGAGAGGCGCATACACTCGAACATTCCTGGGGGAAGAGC 342
 QY 241 TACGGGGAGCAAAACGACATCTGGAGTTCACCCATGACAGTACAGGAGCTGCAGAGG 300
 Db 343 TACGGGGAGCAAAACGACATCTGGAGTTCACCCATGACAGTACAGGAGCTGCAGAGG 402
 QY 301 TACGCCGAGGAGGTTGGGATCTTTCTCACTGCTCTGGCATGGATGAGTGCAGTTGAA 360
 Db 403 TACGCCGAGGAGGTTGGGATCTTTCTCACTGCTCTGGCATGGATGAGTGCAGTTGAA 462
 QY 361 TTCTGTCATGAACCTGATGTTCCATTTTCAAAGTTGATCTGGAGACACTAATATTTT 420
 Db 463 TTCTGTCATGAACCTGATGTTCCATTTTCAAAGTTGATCTGGAGACACTAATATTTT 522
 QY 421 CCTTATCTGGAAAGACAGCCAAAGAGTTCGCCAATGGTATCTCCAGTGGGATGCGAG 480
 Db 523 CCTTATCTGGAAAGACAGCCAAAGAGTTCGCCAATGGTATCTCCAGTGGGATGCGAG 582
 QY 481 TCNATGACACCATCAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCGTCG 540
 Db 593 TCNATGACACCATCAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCGTCG 642
 QY 541 TTCTTGAGTGTACAGGCGCATACCGCTCCAGCTGAGGAGCTCAACCTCGGGGCTCATC 600
 Db 643 TTCTTGAGTGTACAGGCGCATACCGCTCCAGCTGAGGAGCTCAACCTCGGGGCTCATC 702
 QY 601 TCGGAATATCAGAGCTCTTCTGACATTCCTAGAGTTCCTAGGATTCCTGGGATGAAACAGCG 660
 Db 703 TCGGAATATCAGAGCTCTTCTGACATTCCTAGAGTTCCTAGGATTCCTGGGATGAAACAGCG 762
 QY 661 ATAGCGATATCTGGCGCGAGTGTCTTGGGGCCCAAGTCTTGGAGAGTCAACATACT 720
 Db 763 ATAGCGATATCTGAGCGCGAGTGTCTTGGGGCCCAAGTCTTGGAGAGTCAACATACT 822
 QY 721 TTGACAAAGCCTGGAAGGGAGTGACCACTCGGCTCTGCTGAGCCTGGAGAACTGGCC 780
 Db 823 TTGACAAAGCCTGGAAGGGAGTGACCACTCGGCTCTGCTGAGCCTGGAGAACTGGCC 881
 QY 781 GAGCTGTGCGGTGAGTGGTCTTGTGAGGTGCTGCTGGGCTGCTGAGGCTGCTGAGGAGCTG 840
 Db 882 GAGCTGTGCGGTGAGTGGTCTTGTGAGGTGCTGCTGGGCTGCTGAGGCTGCTGAGGAGCTG 941
 QY 841 CTGCGCTGTGAGTGGTCTGCAATCAGAGAG 870
 Db 942 CTGCGCTGTGAGTGGTCTGCAATCAGAGAG 970

RESULT 12
 LOCUS AL526241/c 989 bp mRNA linear EST 23-MAY-2003
 DEFINITION AL526241 Homo sapiens NEUROBLASTOMA COX 25-NORMALIZED Homo sapiens
 cDNA clone CS0DC016Y22 3-PRIME, mRNA sequence.
 AL526241
 ACCESSION AL526241.2 GI:31064102
 VERSION 1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 989)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 TITLE

JOURNAL
COMMENT

Unpublished (2001)
 On Feb 13, 2001 this sequence version replaced gi:12789734.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 5047.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DC016BGL1NP1&cluster=5047.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1680
 Faraday Avenue Genoscope sequence ID : CS0DC016BGL1NP1.

FEATURES
source

1..989
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DC016Y22"
 /tissue_type="NEUROBLASTOMA COX 25-NORMALIZED"
 /clone_lib="Homo sapiens NEUROBLASTOMA COX 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 76.9%; Score 830; DB 9; Length 989;
 Best Local Similarity 93.0%; Pred. NC. 3.9e-204; Indels 6; Gaps 4;
 Matches 872; Conservative 34; Mismatches 26;
 QY 126 GCGCCAGGAGTGTGGGGCTGATTTGTGCCAAGTTCCAGAGAGTGTAGATTCAGTT 185
 Db 94C KCCCAAGGAGTGTGGGGCTGATGTG---CTAAGTCCAGAGAGTGTAGATTCAGTT 884
 QY 186 TAATCGGAAGCCTTGGAGAGGCGCATACACTCTGGAAGCATTCCTGGGGGAAGAGCTACCG 245
 Db 883 TAATCGGAAGCCTTGGAGAGGCGCATACACTCTGGAAGCATTCCTGGGGGAAGAGCTACCG 824
 QY 246 GGAGCACAAAACACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGACAGGAGTACCG 305
 Db 823 GGAGCACAAAACACATCTGGAGTTCAGCCATGACCAAGTACAGGAGCTGACAGGAGTACCG 764
 QY 306 CGAGGAGTGTGGATCTTCTTCACTGCTCTGGCATGGATGAGATGGCATTCGATTCCT 365
 Db 763 CGAGGAGTGTGGATCTTCTTCACTGCTCTGGCATGGATGAGATGGCATTCGATTCCT 704
 QY 366 GCATCAACTGAATGTTCATTTTCAAGTGGATCTGGAGACATTAATATTTTCTTA 425
 Db 703 GCATCAACTGAATGTTCATTTTCAAGTGGATCTGGAGACATTAATATTTTCTTA 644
 QY 426 TCTGGAAGAGCAGCAAAAAGAGTTCGCCAATGGTGTATCTCCAGTGGGATGCAGTCAAT 485
 Db 643 TTTGGAAGAGCAGCAAAAAGAGTTCGCCAATGGTGTATCTCCAGTGGGATGCAGTCAAT 584
 QY 486 GGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGCTTCT 545
 Db 583 GGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCACTTCTGCTTCT 524
 QY 546 GCAGTGTACAGCGCATACCGCTCTCAGCTGAGAGAGCTCAACCTGCGGGTCACTTCGGA 605
 Db 523 GCAGTGTACAGCGCATACCGCTCTCAGCTGAGAGAGCTCAACCTGCGGGTCACTTCGGA 465
 QY 606 ATATCAGAGCTCTTTCTGTCATTTCCCATGGGTATTTCTGGGCATGAGACGGCATAGC 665
 Db 464 ATATCAGAGCTCTTTCTGTCATTTCCCATGGGTATTTCTGGGCATGAGACGGCATAGC 405
 QY 666 GATATCTGTGCGCGCAGTGGCTCTGGGGCCCAAGGTGTTGGAACCTCACAATCTTTGGA 725
 Db 404 GATATCTGTGCGCGCAGTGGCTCTGGGGCCCAAGGTGTTGGAACCTCACAATCTTTGGA 345
 QY 726 CAAGACCTGGAAGGGAGTG-ACCACTCGGCTCTGAGCCTGGAACCTGGCCGAGC 784


```

JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgaps-remail.nih.gov
              Tissue Procurement: ATCC/DCTD/DTF
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM13530 row: e column: 10
              High quality sequence stop: 711.
              Location/Qualifiers
              1..872
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:6167793"
              /tissue_type="melanotic melanoma"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
              Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
              Average insert size 2 kb. Library constructed by Life
              Technologies."

ORIGIN
Query Match      74.4%; Score 803.2; DB 13; Length 872;
Best Local Similarity 99.0%; Pred. No. 3.3e-197;
Matches 808; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATCCGCTGAGCTGAGCTGTGTCCCGGCGCTGGTGGGGGCGGCAACACCGTGTCTC 60
DB 57 ATCCGCTGAGCTGAGCTGTGTCCCGGCGCTGGTGGGGGCGGCAACACCGTGTCTC 116
QY 61 ATCATTTCCGAGATCGCCAGAACCCACGAGGCGACCTGACGTAGCCAGCGCATGATC 120
DB 117 ATCATTTCCGAGATCGCCAGAACCCACGAGGCGACCTGACGTAGCCAGCGCATGATC 176
QY 121 CGCATGCCAAGGAGTGTGGGCTGATTGTGCCAAGTTCAGAGAGAGTGAGCTAGAAATTC 180
DB 177 CGCATGCCAAGGAGTGTGGGCTGATTGTGCCAAGTTCAGAGAGAGTGAGCTAGAAATTC 236
QY 181 AAGTTTAATCGAAGCCTTGGAGAGGCCATACACCTCGAAGCATTTCTGGGGGAGAGC 240
DB 237 AAGTTTAATCGAAGCCTTGGAGAGGCCATACACCTCGAAGCATTTCTGGGGGAGAGC 296
QY 241 TAGCGGGAGCACAAACGACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 300
DB 297 TAGCGGGAGCACAAACGACATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCAGAGG 356
QY 301 TAGCGGGAGGAGTGGGATCTTCTTCACTGCTCTGCGATGGATGAGATGSCAGTTGAA 360
DB 357 TAGCGGGAGGAGTGGGATCTTCTTCACTGCTCTGCGATGGATGAGATGSCAGTTGAA 416
QY 361 TTCTGCTGATCACTGAATGTTCATTTTCAAAGTTGGAATCGAGAGACACTAATAATTTT 420
DB 417 TTCTGCTGATCACTGAATGTTCATTTTCAAAGTTGGAATCGAGAGACACTAATAATTTT 476
QY 421 CCTTATCTGAAAGACAGACGACCAAAAGGTCGCCAATGGTGTGATCTCCAGTGGGATCAG 480
DB 477 CCTTATCTGAAAGACAGACGACCAAAAGGTCGCCAATGGTGTGATCTCCAGTGGGATCAG 536
QY 481 TCAATGACACCATGAAGAGCAAGTTTATCAGATCGTGAAGCCCTTCAACCCCAACTTCTGC 540
DB 537 TCAATGACACCATGAAGAGCAAGTTTATCAGATCGTGAAGCCCTTCAACCCCAACTTCTGC 596
QY 541 TTCTTGAGTGTACACGCGATACCCCGCTCAGCTGAGGAGCTCAACCTCGCGGCTCATC 600
DB 597 TTCTTGAGTGTACACGCGATACCCCGCTCAGCTGAGGAGCTCAACCTCGCGGCTCATC 656
QY 601 TCGGAATATCAGAGCTCTTTTCTTGACATTCCTCATAGGATTTCTGGGCGATGAACAGGC 660

```

```

DB 557 TCGGATATCAGAAGCTCTTTCCTGACATTCCTCCATAGGATTCCTGGGCGATGAACAGC 716
QY 661 ATACCGATATCTGTGGCGCAGTGGCTCTGTGGGGGCAAGGTGTGGGAACGTACATAACT 720
DB 717 ATACCGATATCTGTGGCGCAGTGGCTCTGTGGGGGCAAGGTGTGGGAACGTACATAACT 776
QY 721 TTGACACAGACCTGGAAGGGAGTGAACACTTCGGCCTCGCTGGAGCTGGAGACTGGCC 780
DB 777 TTGACACAGACTGGAAAGGGAGTGAACACTTCGGCCTCGCTGGAGCTGGAGACTGGCC 836
QY 781 GAGCTGTGTGGCTCAGTGGCTCTTGTGTGGAGCGTGC 816
DB 837 GAGCTGTGTGGCTCAGTGGCTCTTGTGTGGAGCGTGC 872

RESULT 15
BQ064521 1031 bp mRNA linear EST 02-APR-2002
LOCUS AGENCOURT_6853542 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5926258
5', mRNA sequence.
DEFINITION BQ064521.1 GI:19893231
ACCESSION BQ064521
VERSION BQ064521.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1031)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
          Email: cgaps-remail.nih.gov
          Tissue Procurement: Lou Staudt
          cDNA Library Preparation: Rubin Laboratory
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCV2099 row: e column: 11
          High quality sequence start: 22
          High quality sequence stop: 736.
          Location/Qualifiers
          1..1031
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="IMAGE:5926258"
          /tissue_type="lymphoma, cell line"
          /lab_host="DH10B (phage-resistant)"
          /clone_lib="NIH_MGC_99"
          /note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:
          EcoRI; cDNA made by oligo-dT priming. Directionally cloned
          into EcoRI/XhoI sites using the following 5' adaptor:
          GGACGAG(G). Size-selected >500bp for average insert size
          1.8kb. Library constructed by Ling Hong in the laboratory
          of Gerald M. Rubin (University of California, Berkeley)
          using ZAP-cDNA synthesis kit (Stratagene) and Superscript
          II RT (Life Technologies). Note: this is a NIH_MGC
          Library."

ORIGIN
Query Match      73.7%; Score 796.4; DB 13; Length 1031;
Best Local Similarity 99.1%; Pred. No. 2.1e-195;
Matches 811; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 264 GGAGTTCAGGCATGACCGATACAGGAGCTGCAGAGTACCGCAGGAGGTGGATCTT 323
DB 23 GGAGTTCAGGCATGACCGATACAGGAGCTGCAGAGTACCGCAGGAGGTGGATCTT 82
QY 324 CTTCACTGCTCTGGCATGGATGAGATGSCAGTTGAAATTCCTGCGATGAATGTTCC 383

```

Db 83 CTTCACTGCTCTGGGATGGATGAGATGGCAGTGTGAATCTCTCCATGAACTGAATGTTCC 142
QY 384 ATTTTCAAAGTTGGATCTGGAGACACTAATAATTTTCCTTATCTGSAAGACAGCCAA 443
Db 143 ATTTTCAAAGTTGGATCTGGAGACACTAATAATTTTCCTTATCTGSAAGACAGCCAA 202
QY 444 ABAAGTCCGCCAATGGTGTCTCCAGTGGGATGCAAGTCAATGGACACCATGAAGCAAGT 503
Db 203 ABAAGTCCGCCAATGGTGTCTCCAGTGGGATGCAAGTCAATGGACACCATGAAGCAAGT 262
QY 504 TTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGCTTCTTGCAGTGTACCAAGCCGATA 563
Db 263 TTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGCTTCTTGCAGTGTACCAAGCCGATA 322
QY 564 CCGGCTCCAGCTGAGGAGCTCAACCTGCGGGTCAATCTCGGAATATCAGAACTCTTTCC 623
Db 323 CCGGCTCCAGCTGAGGAGCTCAACCTGCGGGTCAATCTCGGAATATCAGAACTCTTTCC 382
QY 624 TGACATTTCCCATAGGGTATTCTGGGCATGAACAGGCATAGCGATATCTGTGCGCGAGT 683
Db 383 TGACATTTCCCATAGGGTATTCTGGGCATGAACAGGCATAGCGATATCTGTGCGCGAGT 442
QY 684 GGCCTGCGGGCCAAAGTGTGGAACTCACATACTTTGGACAAGACCTGSAAGGGGAG 743
Db 443 GGCCTGCGGGCCAAAGTGTGGAACTCACATACTTTGGACAAGACCTGSAAGGGGAG 502
QY 744 TGACCACTCGGCTCGCTGGAGCTCGGAACTGGCGAGCTGGTGGGTCACTGGTCT 803
Db 503 TGACCACTCGGCTCGCTGGAGCTCGGAACTGGCGAGCTGGTGGGTCACTGGTCT 562
QY 804 TGTGGAGCGTGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCT 863
Db 563 TGTGGAGCGTGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCT 622
QY 864 TGAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAATTCGGGAAGGCAACATTCTAAC 923
Db 523 TGAGAAGCTGGGCAAGTCTGTGGTGGCCAAAGTGAAATTCGGGAAGGCAACATTCTAAC 682
QY 924 AATGGACATGCTCAGCGTGAAGTGGGTGAGCCCAAGCTATCTCTCTGAAGACATCTT 983
Db 683 AATGGACATGCTCAGCGTGAAGTGGGTGAGCCCAAGCTATCTCTCTGAAGACATCTT 742
QY 984 TAATCTAGTGGGCAAGAGTCTGTGCTCACTGTGTGAAGAGGATGACACCATCATGG-AAG 1042
Db 743 TAATCTAGTGGGCAAGAGTCTGTGCTCACTGTGTGAAGAGGATGACACCATCATGGAG 802
QY 1043 AATTGGTAGATTAATCATGGCAAAAAAATCAAGTCTTAA 1080
Db 803 AATTGGTAGATTAATCATGGCAAAAAAATCAAGTCTTAA 840

Search completed: May 11, 2004, 11:46:12
Job time : 3140.76 secs

361 TTCCTGCATGAAGTGAAGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 420
Db |||||
361 TTCCTGCATGAAGTGAAGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 420
Qy |||||
421 CTTTATCTGGAAAAGACAGCCAAAAGGTGCGCCAAATGCTGATCTCCAGTGGGATGAG 480
Db |||||
421 CTTTATCTGGAAAAGACAGCCAAAAGGTGCGCCAAATGCTGATCTCCAGTGGGATGAG 480
Qy |||||
481 TCAATGGACACCATGAAGCAGTGTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540
Db |||||
481 TCAATGGACACCATGAAGCAGTGTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540
Qy |||||
541 TTCCTGCATGAAGTGAAGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 600
Db |||||
541 TTCCTGCATGAAGTGAAGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 600
Qy |||||
601 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGCTATCTGGGATGAACACAGGC 660
Db |||||
601 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGCTATCTGGGATGAACACAGGC 660
Qy |||||
661 ATAGGATATCTGTGGCGCAGTGGCTCTGGGGGCGCAAGTGTGGAAGTCACTAAT 720
Db |||||
661 ATAGGATATCTGTGGCGCAGTGGCTCTGGGGGCGCAAGTGTGGAAGTCACTAAT 720
Qy |||||
721 TTGGACAAAGCTGGAAGGGAGTGAACCTCGGCTCGCTGGAGCTGGAGAACTGGGC 780
Db |||||
721 TTGGACAAAGCTGGAAGGGAGTGAACCTCGGCTCGCTGGAGCTGGAGAACTGGGC 780
Qy |||||
781 GAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840
Db |||||
781 GAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840
Qy |||||
841 CTGGCTGTGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Db |||||
841 CTGGCTGTGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Qy |||||
901 ATTCCGGAAGCACCATTCTAAATGAGACATGCTCAGCGTGAAGTGGTGGTGGTGGTGG 960
Db |||||
901 ATTCCGGAAGCACCATTCTAAATGAGACATGCTCAGCGTGAAGTGGTGGTGGTGGTGG 960
Qy |||||
961 GCCTATCTCTCTGAGACATCTTTTAACTAGTGGGCGAAGAGTCTGTGCTCTGTTGAA 1020
Db |||||
961 GCCTATCTCTCTGAGACATCTTTTAACTAGTGGGCGAAGAGTCTGTGCTCTGTTGAA 1020
Qy |||||
1021 GAGGATGACACCATCATCGAAGAAATTTGGTAGATAATCATGGCAAAATCAAGTCTTAA 1080
Db |||||
1021 GAGGATGACACCATCATCGAAGAAATTTGGTAGATAATCATGGCAAAATCAAGTCTTAA 1080

RESULT 2

US-09-620-312D-1035
; Sequence 1035, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Fang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Exmanac, Radoie T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and

; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: PC FL-Genes Version 1.0
; SEQ ID NO 1035
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (119)..(1198)
; US-09-620-312D-1035

Query Match 99.7%; Score 1076.8; DB 4; Length 1230;
Best Local Similarity 99.8%; Pred. No. 1.1e-301;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGCGCTGAGCTGGAGCTGTGTCCTGGGGGCGTGGGCTGGGGGCGCAACACCCGCTGTC 60
Db 119 ATGCGCTGAGCTGGAGCTGTGTCCTGGGGGCGTGGGCTGGGGGCGCAACACCCGCTGTC 178
Qy 61 ATCATTTCCGAGATGCGCCAGAACACACAGGGGAGCTGGAGCTAGCCAGCCATGATC 120
Db 179 ATCATTTCCGAGATGCGCCAGAACACACAGGGGAGCTGGAGCTAGCCAGCCATGATC 238
Qy 121 CCATGGCCCAAGAGTGTGGGCTGATTCCTGCCAGTTCAGAGAGTGGAGTGGATTC 180
Db 239 CCATGGCCCAAGAGTGTGGGCTGATTCCTGCCAGTTCAGAGAGTGGAGTGGATTC 298
Qy 181 AAGTTTAAATCGAAAGCCCTTGGAGAGCCATACACCTCGAAAGCTTCTGGGGGAAAGAG 240
Db 299 AAGTTTAAATCGAAAGCCCTTGGAGAGCCATACACCTCGAAAGCTTCTGGGGGAAAGAG 358
Qy 241 TACGGGGAGACAAAGAGCATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCGAGG 300
Db 359 TACGGGGAGACAAAGAGCATCTGGAGTTCAGCCATGACAGTACAGGAGCTGCGAGG 418
Qy 301 TACGGCCGAGGAGTGGGATCTTCTCACTGCTCTGGCATGGATGGATGGAGTTGAA 360
Db 419 TACGGCCGAGGAGTGGGATCTTCTCACTGCTCTGGCATGGATGGATGGAGTTGAA 478
Qy 361 TTCCTGCATGAAGTGAAGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 420
Db 479 TTCCTGCATGAAGTGAAGTTCATTTTCAAAAGTTGGATCTGGAGACACTAATAATTTT 538
Qy 421 CTTTATCTGGAAAAGACAGCCAAAAGGTGCGCCAAATGCTGAGAGCCCTCAACCCCACTTCTGC 540
Db 539 CTTTATCTGGAAAAGACAGCCAAAAGGTGCGCCAAATGCTGAGAGCCCTCAACCCCACTTCTGC 598
Qy 481 TCAATGGACACCATGAAGCAGTGTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540
Db 599 TCAATGGACACCATGAAGCAGTGTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 658
Qy 541 TTTTGGAGTGAACAGGCGATACCCGCTCCAGCTGAGGAGCGTCAACCTGGGGTCAATC 600
Db 659 TTTTGGAGTGAACAGGCGATACCCGCTCCAGCTGAGGAGCGTCAACCTGGGGTCAATC 718
Qy 601 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGCTATCTGGGATGAACACAGGC 660
Db 719 TCGGAATATCAGAACTCTTTCTGATCATTCATAGGCTATCTGGGATGAACACAGGC 778
Qy 661 ATAGGATATCTGTGGCGCAGTGGCTCTGGGGGCGCAAGTGTGGAAGTCACTAAT 720
Db 779 ATAGGATATCTGTGGCGCAGTGGCTCTGGGGGCGCAAGTGTGGAAGTCACTAAT 838
Qy 721 TTGGACAAAGCTGGAAGGGAGTGAACCTCGGCTCGCTGGAGCTGGAGAACTGGGC 780

839 TTGACCAAGACCTGGAAGGGGAGTGAACCACTCGGCTCGCTGGAGCTGGAGAACTGGCC 898
781 GAGCTGGTGGCGTCAGTGCCTGTTGTGGAGCTGCTGCCCTGGCTGCCCAACCAAGCAGCTG 840
899 GAGCTGGTGGCGTCAGTGCCTGTTGTGGAGCTGCTGCCCTGGCTGCCCAACCAAGCAGCTG 958
841 CTGGCTGTGAGATGGCTGCAATGAGAGCTGGCAAGCTGTGGTGGCCCAAGTGA 900
959 CTGGCTGTGAGATGGCTGCAATGAGAGCTGGCAAGCTGTGGTGGCCCAAGTGA 1018
901 ATTCGGAAGGACCACTTCAACATGACATGCTACCGCTGAAAGTGGTGGCCCAAG 960
1019 ATTCGGAAGGACCACTTCAACATGACATGCTACCGCTGAAAGTGGTGGCCCAAG 1078
961 GCCTATCTCTCTGAAGACATCTTAACTAGTGGCAAGAGTCTCTGGTCACTGTGAA 1020
1079 GGCTATCTCTCTGAAGACATCTTAACTAGTGGCAAGAGTCTCTGGTCACTGTGAA 1138
1021 GAGATGACACCATCATGGAAGAAATTTGTTAGATATCATGCAAAATAAAGTCTTAA 1080
1139 GAGATGACACCATCATGGAAGAAATTTGTTAGATATCATGCAAAATAAAGTCTTAA 1198

RESULT 3
US-08-916-421B-1
Sequence 1, Application US/38916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bull et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)
OTHER INFORMATION: n equals a, t, c, or g

```
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664855)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
Query Match
Best Local Similarity 6.8%; Score 73.6; DB 4; Length 1664976;
Matches 300; Conservative 0; Mismatches 334; Indels 6; Gaps 2;
QY 275 ATGACCACTACAGGAGCTGACAGAGTACGCGGAGAGTTGGGATCTTCTCACTGCT 334
DB 465673 AGGAGCAATCTATGAACCTAAAGAAATATGCAAGAAATTTGAATTAATCTCA 465732
QY 335 CTGGATCGATGAGTGCAGTGTGATCTTCATGAATCTGATGATCTTCAAG 394
DB 465733 CACCAATTAATTAATGATGATTAATTAATTAATTAATTAATTAATTA 465792
QY 395 TTGGATCTGGACACTAATAATTTCTTATCTGAAAGACACCCAAAAGTGCGC 454
DB 465793 TTGCTCTGGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 465852
QY 455 CAATGGTGATCTCCAGTGGGATGAGTCAAT---GGACACCACTGAAGCAAGTTATCAGA 511
```

RESULT 4

```
US-08-976-259-81
; Sequence 81, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740032/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
```

TOPOLOGY: linear
US-08-976-259-81

Query Match 4.9%; Score 52.6; DB 4; Length 1234;
Best Local Similarity 48.5%; Pred. No. 2.9e-05;
Matches 179; Conservative 0; Mismatches 184; Indels 6; Gaps 1;

QY 62 TCATTGCGGATGGCCAGAACACACAGGCGGACCTGACGATGACGAGCGATGATCC 121
Db TGATGCGAGATTAATCACTTGGTGATATGACCGGATGAATGATCAATGATG 595
QY 122 GCATGGCCAGAGGTGGGGTGTATGGCCAGTCCAGAGAGTGGAGTGAATTC 181
Db TAGCGGCCACAGCGAGGCTGACTATCTCAACTCGAGAGCGTGTATGTAAGT 655
QY 182 AGTTAATCGAAGCCCTTGAGAGGCCATACACCTCGAAGCATTCCTGGGGAGAGCGT 241
Db TCTATAGCAGGAGAGCTGGAGTCACCGTACCACTC-----TCCCTTTGGCACCACT 709
QY 242 ACGGGAGCACAACGACATCTGGAGTTCAGCCATGACGATGACGAGGAGCTGCAAGGT 301
Db TTAGGGCATCTGGCATGGCATGAACTCAATGAAGAGCAATTTCTCTTGTGCACTCT 769
QY 302 ACGCGGAGGTTGGGATCTCTCACTGCTCTGGCATGATGATGAGATGGCAGTTGAAT 361
Db TCTGTAAGAGATTGGTATCGCTGGTGTCTTATTTAGATATGCCCTGCTGATGAGT 829
QY 362 TCCTGATGAATGAATTTCCATTTTCAAGTTGGATCTGGAGACACTAATAATTTTC 421
Db TCATTGCGCAATTTGAACAGATATGATCAAGCTACCATCAACTATATCTGAACATAAG 889
QY 422 CTTATCTGG 430
Db 890 ATTATTTGG 898

RESULT 5

US-09-495-406-1
Sequence 1, Application US/09495406
Patent No. 6503744

GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarichuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
FILE REFERENCE: 019633-000110US
CURRENT APPLICATION NUMBER: US/09/495,406
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
NUMBER OF SEQ ID NOS: 35
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 1
LENGTH: 11474
TYPE: DNA
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
OTHER INFORMATION: including LOS biosynthesis locus
US-09-495-406-1

Query Match 4.1%; Score 44.4; DB 4; Length 11474;
Best Local Similarity 47.2%; Pred. No. 3.022;
Matches 169; Conservative 0; Mismatches 186; Indels 3; Gaps 1;

QY 363 CQTSCATGAATCAATGTTCCATTTTCAAGTTGGATCTGGAGACACTAATATTTC 422
Db CTTAGAGATATGGGAGTTAGTCTTTTAAAGTTGGTCTAGGTTAGTGAATTAATATCC 7330
QY 423 TTATCTGGAACACAGCGCCAAAGAGGTGCGCCCAATGGTGATCTCCAGTGGGATCGAGTC 482
Db GCTTATTAACACATAGCAGCGCTTTAAAGCCCTTAATGATAGTAGTACACAGGATGATAG 7390

QY 483 AATGGACACCATGAAGCAAGTTTATCAGATCGTGAAGCCCTCAACCCCAACTTCTGCTT 542
Db TATTGAAGATATAAACAACCAACTGTAAATATCTTATTAGACATGAATTCCTTTTCTTTT 7450
QY 543 CTTGAGTGTACAGCGGATACCCGCTCAGCTGAGGACGTCACCTCGGGGTATCTC 602
Db AATGCACACACCAATCTTTTACCCAAACCCGATATCTTTTGAAGATTAAACGCTATGCT 7510
QY 603 GGAATATCAGAGCTCTTTCTGACATTCCTGATGAGGTATCTGGGATGAAACAGCAT 662
Db TGAATTAATAAAGAAATTTCTTTATGG---TAGCTTAAGCGACCAACACACAGATTA 7567
QY 663 AGCGATATCTGTGGCGCAGTGGCTCTGGGGCCCAAGGTGTTSSAACGTCACTAACT 720
Db TCTTCGGTGTAGTGGGTGCACTTGGTCTTGTGCTTGAAGACATTTTACT 7625

RESULT 6

US-09-252-991A-6409
Sequence 6409, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6409
LENGTH: 1083
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6409

Query Match 4.0%; Score 43.4; DB 4; Length 1083;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 144; Conservative 0; Mismatches 146; Indels 3; Gaps 1;

QY 26 CCGGCGCTGCTGGTGGGGCAACACCCGCTTTCATCATTCGCGAGATCGGCCAGAAC 85
Db CCAAGCGCGTCTGTGTCATCGCCAGGGCGGAGTACATCAATCAACGCGTGCAGC 482
QY 86 ACAGGGCGGACCTGAGGTAGCCAGCGCATGATCCGATGCGCAAGAGGTGTGGGGCTG 145
Db CCAATGCGCTGCTGTACGGGTGACCCGCT---TACCGCTTCGAGCGAGCGTGGCTGG 539
QY 146 ATTGTGCCAAGTTCACAGAGGTGAGCTAGAAATTCAGTTTAAATCGGAAGCCCTTGAGA 205
Db ATGCGCCAGTTTCGCCAGCGCGCTTCGAGACCGATCACTGGCACTGGAAGAGG 599
QY 206 GGCCATACACTCGAAGCATTTCTGGGGGAGAGAGTACGGGGAGCAACAAACGACATCGG 265
Db TCACCACCACTTGTCTGATCCGAGGGAAGCGTTCCGAGTGGTCAAGTACCGACCG 659
QY 266 AGTTCAGCCATGACAGTACAGGAGCTGCGAGGTTACGCGGAGAGGTGGG 318
Db AACGCTGGGATGCCCACTCAGCCCGCACTGCTGAACACTGTGTGTATGAGG 712

RESULT 7

US-09-252-991A-6241
Sequence 6241, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6241
LENGTH: 1104
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6241

Query Match 4.0%; Score 43.4; DB 4; Length 1104;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 144; Conservative 0; Mismatches 146; Indels 3; Gaps 1;
Qy 26 CCAGGCGCTGGTGGGCGGGAACACCCGGTCTTCATATGCGGAGATCGCCAGAAC 85
Db 467 CCAAGCGCGTCTGTGGCATCGCCAGGCGCGAGTACATCCACATCAACGCGTGCAGC 526
Qy 86 ACCAGGCGACCTGGAGTACGCAAGCGCATGATCCGATGCGCAAGAGTGTGGCGTG 145
Db 527 CCATGGCGTGTGTACGGCTGACCCGCG---TACCGCTTCGACGACGACGGTGGCTGG 583
Qy 146 ATTGTGCCAAGTTCACAGAGAGTGGCTAGAAATTCAGATTAAATCGGAAGCCCTGGAGA 205
Db 584 AATCGGCCAGTTTCGCCAAGCGCGCGCTTCGAGACCGATCATCTGCAACTGGAAGG 643
Qy 206 GGCCATACCTCGAAGCATTCTGGGGAGAGAGTACGGGAGACAACACGACATCTGG 265
Db 644 TCACCCACCTCTGCTGATCGAGGGAAGCGTCCGAGGTGGTCAAGTACCGACCG 703
Qy 266 AGTTAGCCATGACCATGACGAGGAGTGCAGAGGATCGGAGGATCGCGAGGAGTTGG 318
Db 704 AACGCTGGGATGCCAGCTCAGCCCGCACTGCTGAACACTGCTGGTGGTATGGAG 756

RESULT 8
US-08-232-463-14/c
Sequence 14, Application US/0822463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, P. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/232,463
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE: EP 91 114 300.5
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgt-Fls
US-08-232-463-14

Query Match 3.8%; Score 41.4; DB 1; Length 7218;
Best Local Similarity 4.5%; Pred. No. 0.13;
Matches 18; Conservative 209; Mismatches 170; Indels 0; Gaps 0;
Qy 58 CAGGGCGACCTGGAGCTAGCCAGCGCATGCCATGGCCAGGAGTGGGGCTGAT 147
Db 1435 CRR 1376
Qy 148 TGTCGAAGTTCAGAGAGTACGCTAGAAATTCAGTTTAAATCGAAAGCCTTGGAGG 207
Db 1375 RRR 1316
Qy 208 CCATACACTCGAAGCATCTCTGGGGAGAGCTAGCGGGAGACAAACGACATCTGGAG 267
Db 1315 RRR 1256
Qy 268 TCAGCCATGACCATGACAGGAGCTCAGAGGTACCGGAGAGTGGGATCTCTTC 327
Db 1255 RRR 1196
Qy 328 ACTGCTCTGCGCATGAGATGCGAGTGAATTCCTGATGAATGAATGTTCCATT 387
Db 1195 RRR 1136
Qy 388 TTCAGAGTGGATCTGGAGACATATAATTTCTTATCTGGAAGACAGCAAAAA 447
Db 1135 RRR 1076
Qy 448 GTCGCGCAATGTGATCTCCAGTGGGATCGACTCAA 484
Db 1075 RRRRRRRATCGAAGCTCTCTGACCTCAGCCAA 1039

RESULT 9
US-09-833-381-1478
Sequence 1478, Application US/0983381
Patent No. 6672186
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
FILE REFERENCE: 5800-119
CURRENT APPLICATION NUMBER: US/09/833,381
CURRENT FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: 09/516,448
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 2050
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1478
LENGTH: 2437
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(2437)
OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1478

Query Match 3.8%; Score 41; DB 4; Length 2437;

Best Local Similarity 52.7%; Pred. No. 0.095;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 724 GACAGACCTGGAAGGAGTGCACCTCGCTCGCTGGAGCTGGAGAACTGGCCGAG 783
Db 1803 GCCAAGCGGAGAGGACCGCAAGGCCATCGCCGCTTCAGAGGAGGACAGCAAGGCGG 1862
QY 784 CTGCTGGGTCAGTGGCTCTTGTGGAGCGTGCCTGGCTCCCAACCAAGCAGCTGCTG 843
Db 1863 CAGGAGCAGCAGGAGGCGCCAGTGGGCGCCAGCCCGAGCGCCAGCACTGCAGAGAAG 1922
QY 844 CCCTGTGAGATGCCCTCAATGAGAAGCTGGGCAAGCTGTGGTGCCA 592
Db 1923 CAGCCTGAGAAGGAGGAGGAGGAGGCGCGGAGAGAGTGTGGAGGACA 1971

RESULT 10

US-09-833-381-1479
; Sequence 1479, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1479
; LENGTH: 2437
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2437)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1479

Query Match 3.8%; Score 41; DB 4; Length 2437;
Best Local Similarity 52.7%; Pred. No. 0.095;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 724 GACAGACCTGGAAGGAGTGCACCTCGCTCGCTGGAGCTGGAGAACTGGCCGAG 783
Db 1803 GCCAAGCGGAGAGGACCGCAAGGCCATCGCCGCTTCAGAGGAGGACAGCAAGGCGG 1862
QY 784 CTGCTGGGTCAGTGGCTCTTGTGGAGCGTGCCTGGCTCCCAACCAAGCAGCTGCTG 843
Db 1863 CAGGAGCAGCAGGAGGCGCCAGTGGGCGCCAGCCCGAGCGCCAGCACTGCAGAGAAG 1922
QY 844 CCCTGTGAGATGCCCTCAATGAGAAGCTGGGCAAGCTGTGGTGCCA 592
Db 1923 CAGCCTGAGAAGGAGGAGGAGGAGGCGCGGAGAGAGTGTGGAGGACA 1971

RESULT 11

US-09-252-991A-13238
; Sequence 13238, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13238
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13238

Query Match 3.7%; Score 40; DB 4; Length 738;
Best Local Similarity 55.9%; Pred. No. 0.1;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 218 CGAAGCATCTCTGGGGAAGACGTACCGGGAGACAAACAGACATCTGAGTTTCAGCCATG 277
Db 317 CGCCGCGCTCTTCAGAAACATCTACAGGAACTCAACCGGACCTGAAATCCCGATTC 376
QY 278 ACCAGTACAGGAGCTGCAGAGGTACCGCGAGGAGTGGAGTCTTCTTCACTGCTCTG 337
Db 377 CCAACAACGGCTACCTGCAGCGCTGGCGGAGCAGGGCGTGTGTGCTCATACCTCC 436
QY 338 GCATGGATGATGGC 353
Db 437 TGACCGTCGAGCAGGC 452

RESULT 12

US-09-252-991A-9714
; Sequence 9714, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9714
; LENGTH: 363
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9714

Query Match 3.5%; Score 38.2; DB 4; Length 363;
Best Local Similarity 49.3%; Pred. No. 0.23;
Matches 100; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 697 AAGTGTGGAAGTGCATTAACCTTTGGACAAGACCTGGAGAGGAGGAGTACCACTCGGCC 756
Db 19 ATGATGCCCGCAGGCTCGCCGCTGACCTCGAGCGTGTGTGGAGCAACGACCCGCT 78
QY 757 TCGCTGGAGCTCGGAACTGGCCGAGCTGGTGGCTCAGTGGCTCTTGTGGAGCTGCC 816
Db 79 CCGCTGGAGCAGATGCGCTTGACCGAAGCAGCGCTGGAGCAGGCGCAAGCGGTGGGTGCC 138
QY 817 CTGGGCTCCCAACCAAGCAGCTGCTGCCCTGTGAGATGGCTGTCAATGAAGAAGCTGGC 876
Db 139 ACCGACGACCTGGCCGAGCTGAACTGGCCAGCAAGTACGCGCGGCGGCGAGATCGG 198
QY 877 AAGTGTGCTGGCCAAAGTAA 899
Db 199 ATGACCGCGGAGTCTTACAGAA 221

RESULT 13

US-09-252-991A-9674
; Sequence 9674, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9674
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9674

Query Match 3.5%; Score 38.2; DB 4; Length 1254;
Best Local Similarity 49.3%; Pred. No. 0.43;
Matches 100; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 697 AAGGTGTGGAACTGACATTAACCTTGGACAAAGACCTGGAAGGGGAGTGACCACTCGGCC 756
DB 101 ATGATGGCCGAGGCTCGCCGCGCTGACCTGAGCGGTGTGTCGAGCAACGACCGGCT 160
QY 757 TCGTGGAGCTGGAGAACTGCGGAGCTGTGGGCTGAGTGCCTGCTTGTGGAGCGTCCC 816
DB 161 CCGCTGGAGCAGATGCGGCTGACCGAAACAGACGCGCTGGAGCAGCGCCAAAGCGGTGGTCCC 220
QY 817 CTGGGCTCCCAACCAACAGCACTGCTGCTGTGAGATGGCTGCAATGAGAACTGGGC 876
DB 221 ACCGACAGCTGGCGGAGCTGAACCTGGCCCGAGCAAGTACGCGCGCGCGATCGCG 280
QY 877 AAGTCTGTGTGGCGCAAGTGAA 899
DB 281 ATGACCGCGAGTCTCTACAAGAA 303

RESULT 14
US-09-252-991A-9533/c
; Sequence 9533, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9533
; LENGTH: 3654
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9533

Query Match 3.5%; Score 38.2; DB 4; Length 3654;
Best Local Similarity 49.3%; Pred. No. 0.75;
Matches 100; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 697 AAGGTGTGGAACTGACATTAACCTTGGACAAAGACCTGGAAGGGGAGTGACCACTCGGCC 756
DB 550 ATGATGGCCGAGGCTCGCCGCGCTGACCTGAGCGGTGTGCGAGCAACGACCGGCT 491
QY 757 TCGTGGAGCTGGAGAACTGCGGAGCTGTGGGCTGAGTGCCTTGTGGAGCGTCCC 816
DB 490 CCGCTGGAGCAGATGCGGCTGACCGAAACAGGCGCTGGAGCAGCGCCAAAGCGGTGGTCCC 431
QY 817 CTGGGCTCCCAACCAACAGCACTGCTGCTGTGAGATGGCTGCAATGAGAACTGGGC 876
DB 430 ACCGACAGCTGGCGGAGCTGAACCTGGCCCGAGCAAGTACGCGCGCGCGAGATCGCG 371

QY 877 AAGTCTGTGTGGCGCAAGTGAA 899
DB 370 ATGACCGCGAGTCTCTACAAGAA 348

RESULT 15
US-09-252-991A-13452
; Sequence 13452, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13452
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13452

Query Match 3.4%; Score 36.4; DB 4; Length 486;
Best Local Similarity 56.8%; Pred. No. 2.89;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 196 GCCTTGGAGAGGCCATACACCTCGAAGCATTCCTGGGGGAAGACGTACGGGGAGCAAAA 255
DB 74 GCCTATGTGCGCCGAGCGGAGTCTGTGCGGGGTGGCGCTTCCGGAGCAGGC 133
QY 256 CGACATCTGGAGTTTCAGCCATGACCCAGTACAGGGGAGCTGCAGAGGTAGCCCGAGGAGG 313
DB 134 CATCAGCAGCAGCGCGCGGAGGCGGAGCCAGCACGCGGATCAGCAGATAGGCGCGCAGCAAG 191

Search completed: May 11, 2004, 07:22:56
Job time : 99.111 secs